Lectures on General Relativity and Cosmology

Pedro G. Ferreira
Astrophysics
University of Oxford,
DW Building, Keble Road
Oxford OX1 3RH, UK

last modified May 22, 2015

Preamble

The following set of lectures was designed for the Oxford Undergraduate course and are given during the 3rd year of the BA or MPhys course. I have tried to keep the mathematical jargon to a minimum and ground most of the explanations with physical examples and applications. Yet you will see that, although there is very little emphasis on differential geometry you still have to learn what tensors or covariant derivatives are. So be it.

These notes are not very original and are based on a number of books and lectures. In particular I have used

- *Gravitation and Cosmology*, Steven Weinberg (Wiley, 1972)
- *Gravity: An Introduction to Einstein's General Relativity*, James B. Hartle (Addison Wesley, 2003)
- *Part II General Relativity*, G W Gibbons (can be found online on `http://www.damtp.cam.ac.uk/research/gr/members/gibbons/partiipublic-2006.pdf`)
- *General Relativity: An introduction for Physicists*, M.P Hobson, G.P.Efstathiou and A.N.Lasenby (CUP, 2006)
- *General Relativity*, Alan Heavens (not publicly available).
- *Theories of Gravity and Cosmology* T. Clifton, P.G. Ferreira and C. Skordis (Physics Reports, 2012)
- *Cosmological Physics* John Peacock (CUP, 1998)

But there are many texts out there that you can consult.

For a start, I will assume that the time coordinate, t and the spatial coordinate, $\vec{x} = (x^1, x^2, x^3)$ can be organized into a 4-vector $(x^0, x^1, x^2, x^3) = (ct, \vec{x})$ where c is the speed of light. Throughout these lecture notes I will use the $(-,+,+,+)$ convention for the metric. This means that the Minkowski metric is a matrix of the form

$$\eta_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{1}$$

You will note that this is the opposite convention to the one used when you were first learning special relativity. In fact you will find that, in general (but not always), books on General Relativity will use the convention we use here, while books on particle physics or quantum field theory will use the opposite convention.

We will be using the convention that Roman labels (like i, j, etc) span 1 to 3 and label spatial vectors while Greek labels (such as α, β, etc) span 0 to 3 and label space-time vectors.

We will also be using the *Einstein summation convention.* This means that, whenever we have a pair of indices which are the same, we must add over them. So for example, when we write

$$ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta$$

we mean

$$ds^2 = \sum_{\alpha=0}^{3} \sum_{\beta=0}^{3} \eta_{\alpha\beta} dx^\alpha dx^\beta$$

Note that the paired indices always appear with one as a superscript (i.e. "up") and the other one as a subscript (i.e. "down").

Finally, I am extremely grateful to Tessa Baker, Tim Clifton, Rosanna Hardwick, Alan Heavens, Anthony Lewis, Ed Macaulay and Patrick Timoney for their help in putting together the lecture notes, exercises and spotting errors.

1 Why General Relativity?

Throughout your degree, you have learnt how almost all the laws of physics are invariant if you transform between inertial reference frames. With Special Relativity you can now write Newton's 2^{nd} law, the conservation of energy and momentum, Maxwell's equations, etc in a way that they are unchanged under the Lorentz transformation. One force stands apart: gravity. Newton's law of gravity, the inverse square law, is manifestly not invariant under the Lorentz transformation.

You can now ask yourself the question: can we write down the laws of physics so they are invariant under *any* transformation? Not only between reference frames with constant velocity but also between *accelerating* reference frames. It turns out that we can and in doing so, we incorporate gravity into the mix. That is what the General Theory of Relativity is about.

2 Newtonian Gravity

In these lectures we will be studying the modern view of gravity. Einstein's theory of space-time is one of the crowning achievements of modern physics and transformed the way we think about the fundamental laws of nature. It superseded a spectacularly successful theory, Newton's theory of gravity. If we are to understand the importance and consequences of Einstein's theory, we need to learn (or revise) the main characteristics of Newton's theory.

Newton's Law of Universal Gravitation for two bodies A and B with masses m_A and m_B at a separation r can be stated in a simplified form as:

$$F = -G\frac{m_A m_B}{r^2}$$

where $G = 6.672 \times 10^{-11}$ m^{3}kg^{-1}s^{-2}. The resulting gravitational acceleration felt by mass m_A is

$$g = -G\frac{m_B}{r^2}$$

We can rewrite the expression in terms of vectors $\mathbf{r} = (x^1, x^2, x^3)$.

$$\mathbf{F} = G m_A m_B \frac{\mathbf{r}_B - \mathbf{r}_A}{r^3}$$

is the force exerted on the mass m_A and the gravitational acceleration

$$\mathbf{F} = m_A \mathbf{g}$$

Consider now a few simple applications. We can use the above expression to work out the gravitational acceleration on the surface of the Earth. If we Taylor expand around the radius of the Earth, $R_\oplus$ we find

$$g = -G\frac{M_\oplus}{r^2} = -G\frac{M_\oplus}{(R_\oplus + h)^2} \simeq -G\frac{M_\oplus}{R_\oplus^2}(1 - 2\frac{h}{R_\oplus})$$

where h is the height above the surface of the Earth and $M_\oplus$ is the mass of the Earth. With $M_\oplus = 5.974 \times 10^{27}$g and $R_\oplus = 6.378 \times 10^8$cm we find a familiar value: $g \simeq 9.8$ms^{-2}.

The power of Newton's theory is that it allows us to describe, with tremendous accuracy, the evolution of the Solar System. To do so, we need to solve the two body problem. We have that the equations of motion are

$$m_i \ddot{\mathbf{r}}_A = G m_A m_B \frac{\mathbf{r}_B - \mathbf{r}_A}{r^3}$$

with $r = |\mathbf{r}| = |\mathbf{r}_A - \mathbf{r}_B|$. We are interested in how $\mathbf{r}$ evolves and we can find its evolution by simplifying this system. Let us first define the total mass

$$M = m_A + m_B$$

and the reduced mass

$$\mu = \frac{m_A m_B}{M}$$

Ignoring the motion of the centre of mass, we have that the Lagrangian for this system will be

$$L = \frac{1}{2}\mu \dot{\mathbf{r}} \cdot \dot{\mathbf{r}} + G\frac{\mu M}{r}$$

It is more convenient to transform to spherical coordinates so

$$\begin{aligned} x^1 &= r\cos(\phi)\sin(\theta) \\ x^2 &= r\sin(\phi)\sin(\theta) \\ x^3 &= r\cos(\theta) \end{aligned}$$

then the Lagrangian becomes

$$L = \frac{1}{2}\mu(\dot{r}^2 + r^2\dot{\theta}^2 + r^2\sin^2(\theta)\dot{\phi}^2) + G\frac{\mu M}{r}$$

The angular parts of the Euler-Lagrange equations are

$$\begin{aligned} \mu\frac{d}{dt}(r^2\dot{\theta}) &= \mu r^2\sin(\theta)\cos(\theta)\dot{\phi}^2 \\ \frac{d}{dt}\left[\mu r^2\sin^2(\theta)\dot{\phi}\right] &= 0 \end{aligned}$$

We can integrate the second equation to give

$$\mu r^2\sin^2(\theta)\dot{\phi} = J_\phi$$

and, multiplying the first equation by $2r^2\dot{\theta}$, we can integrate to find

$$\mu^2 r^4\dot{\theta}^2 = J_\theta^2 - \frac{J_\phi^2}{\sin^2(\theta)}$$

We can now choose a coordinate system such that the orbit lies on the equatorial plane ($\theta = \pi/2$) and $J_\theta = J_\phi = J$ which corresponds to $\dot{\theta} = 0$. We are left with two coordinates, r and ϕ.

Replacing the the angular momentum conservation equation in the radial equation of motion, we have

$$\mu\ddot{r} - \frac{J^2}{\mu r^3} + G\frac{\mu M}{r^2} = 0$$

which we can integrate to give us an expression for the conserved energy

$$\frac{E}{\mu} = \frac{1}{2}\dot{r}^2 + \frac{1}{2}\frac{J^2}{\mu^2 r^2} - G\frac{M}{r} \tag{2}$$

Note that we can define an effective potential energy which is given by

$$V_{eff}(r) = \frac{1}{2}\frac{J^2}{\mu r^2} - G\frac{\mu M}{r}$$

Again, using conservation of energy, we can use

$$J dt = \mu r^2 d\phi$$

to reexpress time derivatives in terms of angular derivatives

$$\frac{d}{dt} = \frac{J}{\mu r^2}\frac{d}{d\phi} \quad \text{and} \quad \frac{d^2}{dt^2} = \frac{J}{\mu r^2}\frac{d}{d\phi}(\frac{J}{\mu r^2}\frac{d}{d\phi})$$

The Euler Lagrange equation then becomes

$$\frac{J}{r^2}\frac{d}{d\phi}\left(\frac{J}{\mu r^2}\frac{dr}{d\phi}\right) - \frac{J^2}{\mu r^3} = -G\frac{\mu M}{r^2}$$

If we change to $u = 1/r$ we have

$$\frac{d^2u}{d\phi^2} + u = \frac{\mu}{J^2}G\mu M \tag{3}$$

This is the simple harmonic oscillator equation with a constant driving force. We can solve this equation to find

$$u = \frac{1}{r} = \frac{G\mu(\mu M)}{J^2}\left[1 + \sqrt{1 + \frac{2EJ^2}{\mu(G\mu M)^2}}\cos(\phi - \phi_0)\right]$$

where E is the conserved energy of the system (and an integration constant) and ϕ_0 is the turning point of the orbit (and the second integration constant). You should recognize this as the equation for a conic with ellipticity e given by

$$e = \sqrt{1 + \frac{2EJ^2}{\mu(G\mu M)^2}}$$

In other words, the trajectories of the two body problem correspond to closed orbits in the form of ellipses.

The Solar System is almost perfectly described in terms of these trajectories. In fact almost all the planets have quasi-circular orbits with eccentricities $e \simeq 1 - 10\%$. Mercury stands out, with $e \simeq 20\%$. Furthermore, Mercury's orbit does not actually close in on itself but precesses at a rate of about 5600 arc seconds per century. This primarily due to the effect of the other planets slowly nudging it around and, again, can be explained with Newtonian mechanics. But since the mid 19^{th} century, it has been known that there is still an un accounted amount of precession, about 43 arc seconds per century. We shall find its origin in later lectures.

Finally, it is convenient to define the *Newtonian potential* or *gravitational potential* in terms of the potential energy, V, of the system:

$$V \equiv m\Phi$$

The gravitational force will then be

$$\mathbf{F} = -\nabla V$$

and the gravitational acceleration is given by

$$\mathbf{g} = -\nabla\Phi$$

The gravitational potential satisfies the Newton Poisson equation

$$\nabla^2\Phi = 4\pi G\rho$$

3 The Equivalence Principle

In the last section, we worked with two equations: Newton's Law of Universal Attraction and Newton's 2^{nd} Law. In both of these equations a "mass" appears and we assumed that they are one and the same. But let us now rewrite them and be explicit about the different types of mass coming in, the inertial mass m_I and the gravitational mass, m_G:

$$\begin{aligned} m_I \mathbf{a} &= \mathbf{F} \\ V &= -G\frac{m_G M_G}{r} \end{aligned}$$

We assumed that

$$m_I = m_G$$

but are they? Their equivalence has been tested to surprising precision with *Lunar Laser Ranging* observations. This involves bouncing a laser pulse off reflecting mirrors sitting on the surface of the moon and measuring its orbit as it sits in the joint gravitational field of both the Earth and the Sun. The distance between the Earth and the Moon is 384,401 km and has been measured with a precision of under a centimetre

The Lunar Ranging experiment can be seen as a type of *Eötvos Experiment* where the Earth and Moon are the test masses sitting in the gravitational field of the Sun. A laboratory based Eötvos Experiment can be constructed in the following way. Consider two masses made of different material attached to either end of a rod. The rod is suspended from a string on the surface of the Earth. Each of the masses will be subjected to two forces: the gravitational pull to the centre and the centrifugal force. Hence the rod will hang at an angle relative to the vertical direction. The rod is free to rotate if there is a difference in the gravitational acceleration between the masses. This can only happen if there is a difference between m_G and m_I.

We can look at the numbers here. Consider the two masses, which have gravitational masses m_{G1} and m_{G2} and inertial masses m_{I1} and m_{I2}. Denote the component of the gravitational force in the direction that makes the rod twist to be g^t and the acceleration of each of the masses to be a_1^t and a_2^t. We then have that

$$\begin{aligned} m_{I1} a_1^t &= m_{G1} g^t \\ m_{I2} a_2^t &= m_{G2} g^t \end{aligned}$$

If the ratios of inertial to gravitational mass are the same for both bodies, then the acceleration will be the same for both. Any difference in the gravitational and inertial masses will lead to a "twist" in the pendulum which can be characterized in terms of a dimensionless paramater:

$$\eta = \frac{a_1^t - a_2^t}{a_1^t + a_2^t} = \frac{\left(\frac{m_{G1}}{m_{I1}} - \frac{m_{G2}}{m_{I2}}\right)}{\left(\frac{m_{G1}}{m_{I1}} + \frac{m_{G2}}{m_{I2}}\right)}$$

Using beryllium and titanium, the current best constraint on η is

$$\eta = \frac{a_1^t - a_2^t}{a_1^t + a_2^t} = (0.3 \pm 1.8) \times 10^{-13}$$

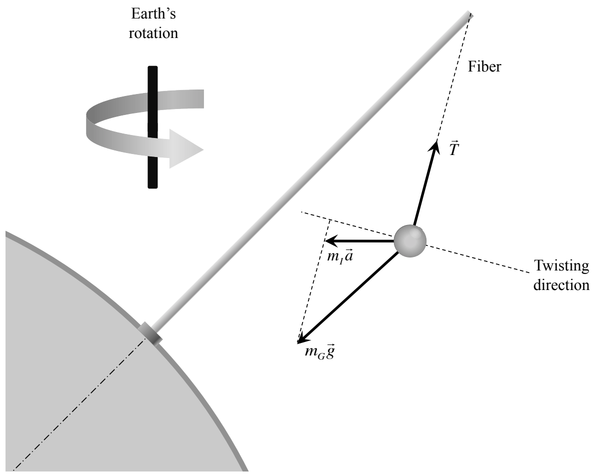

Figure 1: The pendulum for the Eotvos Experiment

which is a factor of 4 better than the original Eotvos experiment of 1922. To date, one of the most successful tests is to use the Earth-Moon system in the gravitational field of the Sun as a giant Eotvos experiment. The difference, with regards to the lab based Eotvos experiments, is that the masses of the test bodies (i.e. the Earth and Moon) are not negligible any more. The test can be done by using lasers reflected off mirrors left on the Moon by the Apollo 11 mission in 1969 and, as mentioned earlier, the constraint is

$$\eta = (-1.0 \pm 1.4) \times 10^{-13}$$

This is one of the most accurately tested principles of physics and we expect this constraint to improve by 5 orders of magnitude when space based tests can be performed.

If the gravitational and inertial masses are the same, then we have that a particle in a gravitational field will obey

$$\ddot{\mathbf{r}} = \frac{m_G}{m_I}\mathbf{g} = \mathbf{g}$$

and this will be true of *any* particle. This has a profound consequence-it means that I can always find a time dependent coordinate transformation (from $\mathbf{r}$ to $\mathbf{R}$),

$$\mathbf{r} = \mathbf{R} + \mathbf{b}(t)$$

such that

$$\ddot{\mathbf{R}} = \mathbf{g} - \ddot{\mathbf{b}}(t) = 0$$

In other words, it is always possible to pick an accelerated reference frame such that the observer doesn't feel the gravitational field at all. For example, consider a particle at rest near the surface of the Earth. It will feel a gravitational pull of $g = -9.8\text{ms}^{-2}$. Now place it in a reference frame such that $b = \frac{1}{2}gt^2$ (such as a freely falling elevator). Then the particle at rest in this reference frame won't feel the gravitational pull.

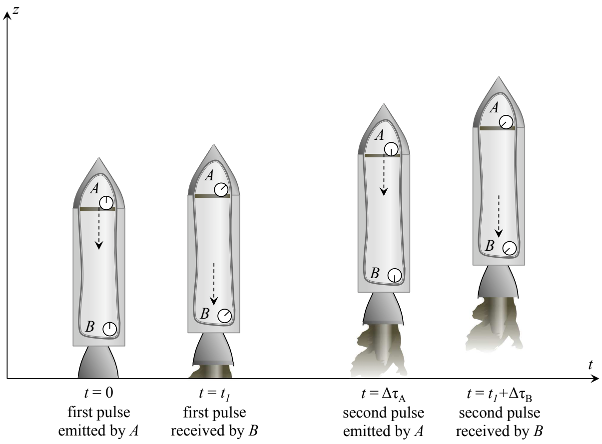

Figure 2: Reference frames for the gravitational redshift

Einstein had an epiphany when he realized this. As he said "*... for an observer falling freely from the roof of a house there exists- at least in his immediate surroundings- no gravitational field*". It led him to formulate the *Equivalence Principle* of which there are three versions that we will state here.

- *Weak Equivalence Principle (WEP)*: All uncharged, freely falling test particles follow the same trajectories, once an initial position and velocity have been prescribed.

- *Einstein Equivalence Principle (EEP)*: The WEP is valid, and furthermore in all freely falling frames one recovers (locally, and up to tidal gravitational forces) the same laws of special relativistic physics, independent of position or velocity.

- *Strong Equivalence Principle (SEP)*: The WEP is valid for massive gravitating objects as well as test particles, and in all freely falling frames one recovers (locally, and up to tidal gravitational forces) the same special relativistic physics, independent of position or velocity.

You will note that these three equivalent principles have different remits. While the first makes a simple statement about the trajectories of freely falling bodies, the second one says something about the laws of physics obeyed by the freely falling bodies and the third addresses going beyond the simplified, point-like test mass approximation. We will use WEP, SEP and EEP interchangeably throughout these lectures although they can be tested in distinct ways.

Let us briefly study one of the consequences of the Equivalence Principle. Consider two observers, one at position A which is at a height h above the surface of the Earth and another at a position B on the surface of the Earth. Observer A emits a pulse every Δt_A to be received by observer B every Δt_B. What is the relation between Δt_A and Δt_B and how is it affected by the gravitational field?

Given what we have seen above, we can think of this as two observers moving upwards with an acceleration g. We have that the positions of A and B are given by

$$\begin{aligned} z_A(t) &= \frac{1}{2}gt^2 + h \\ z_B(t) &= \frac{1}{2}gt^2 \end{aligned}$$

Now assume the first pulse is emitted at $t = 0$ by A and is received at time t_1 by B. A subsequent pulse is then emitted at time Δt_A by A and then received at time $t_1 + \Delta t_B$. We have that

$$\begin{aligned} z_A(0) - z_B(t_1) &= h - \frac{1}{2}gt_1^2 = ct_1 \\ z_A(\Delta t_A) - z_B(t_1 + \Delta t_B) &= h + \frac{1}{2}g\Delta t_A^2 - \frac{1}{2}g(t_1 + \Delta t_B)^2 \\ &\simeq h - \frac{1}{2}gt_1^2 - gt_1\Delta t_B = c(t_1 + \Delta t_B - \Delta t_A) \end{aligned}$$

where we have discarded higher order terms in Δt. Combining the two equations, and assuming $t_1 \simeq h/c$ we have that

$$\Delta t_A \simeq (1 + \frac{gh}{c^2})\Delta t_B$$

In other words, there is gravitational time dilation due to the difference in the gravitational potentials at the two points $\Phi_B - \Phi_A = -gh$ so

$$\text{Rate Received} = \left(1 - \frac{\Phi_B - \Phi_A}{c^2}\right) \times \text{Rate Emitted}$$

This is known as the *gravitational redshift* of light.

4 Geodesics

In the previous section, we saw how important accelerated reference frames could be. From the various equivalence principles, it seems that an accelerating reference frame will be indistinguishable from a frame in a gravitational field. Let us then study how particles and observers travel through space.

In flat space, a particle follows straight lines given by solutions to the kinematic equations of motion

$$\frac{d^2x^\alpha}{d\tau^2} = 0$$

The particle will trace out a path in space-time $x^\alpha(\tau)$ and the solution is of the form

$$x^\alpha(\tau) = x^\alpha(\tau_i) + u^\alpha(\tau - \tau_i)$$

where $x^\alpha(\tau_i)$ and u^α are integration constants.

In the presence of a gravitational field, the path taken by a particle will be curved. Alternatively, in an accelerating reference frame, the same will happen. We call *Geodesics* the shortest paths between two points in space-time. We have just written down the geodesic in a flat, force free space. If we wish to do so in the presence of a gravitational field, we need to solve the *Geodesic equation.* As Einstein argued, according to the principle of equivalence, there should be a *freely falling coordinate system* y^α in which particles move in a straight line and therefore satisfy

$$\frac{d^2y^\alpha}{d\tau^2}=0$$

where τ is the proper time of the particle and hence

$$c^2d\tau^2=-\eta_{\alpha\beta}dy^\alpha dy^\beta$$

Now choose a different coordinate system, x^μ; it can be at rest, accelerating, rotating, etc. We can reexpress the y^αs in terms of the x^μ, $y^\alpha(x^\mu)$. Using the chain rule we have

$$\begin{aligned} 0 &= \frac{d}{d\tau}\left(\frac{\partial y^\alpha}{\partial x^\mu}\frac{dx^\mu}{d\tau}\right) \\ &= \frac{\partial y^\alpha}{\partial x^\mu}\frac{d^2x^\mu}{d\tau^2}+\frac{\partial^2 y^\alpha}{\partial x^\mu\partial x^\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}\end{aligned}$$

Multiplying through by the inverse Jacobian $\partial x^\beta/\partial y^\alpha$ we end up with the *geodesic equation* with the form

$$\frac{d^2x^\beta}{d\tau^2}+\Gamma^\beta{}_{\mu\nu}\frac{dx^\mu}{d\tau}\frac{dx^\nu}{d\tau}=0 \tag{4}$$

where we have defined the *affine connection*

$$\Gamma^\beta{}_{\mu\nu}=\frac{\partial x^\beta}{\partial y^\alpha}\frac{\partial^2 y^\alpha}{\partial x^\mu\partial x^\nu}$$

We can also express the proper time in these new (or arbitrary coordinates) as

$$c^2d\tau^2=-\eta_{\alpha\beta}\frac{\partial y^\alpha}{\partial x^\mu}dx^\mu\frac{\partial y^\beta}{\partial x^\nu}dx^\nu\equiv-g_{\mu\nu}dx^\mu dx^\nu$$

where we have defined the *metric*:

$$g_{\mu\nu}=\eta_{\alpha\beta}\frac{\partial y^\alpha}{\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}$$

The situation is slightly different for a massless particle. Neutrinos or photons follow null paths so $d\tau=0$. Instead of using τ we can use some other parameter σ. We then have

$$\begin{aligned}\frac{d^2y^\alpha}{d\sigma^2} &= 0 \\ -\eta_{\alpha\beta}\frac{dy^\alpha}{d\sigma}\frac{dy^\beta}{d\sigma} &= 0\end{aligned}$$

ne One can repeat the same derivation as above to find

$$\begin{aligned}\frac{d^2x^\mu}{d\sigma^2}+\Gamma^\mu{}_{\alpha\beta}\frac{dx^\alpha}{d\sigma}\frac{dx^\beta}{d\sigma} &= 0\\ -g_{\alpha\beta}\frac{dy^\alpha}{d\sigma}\frac{dy^\beta}{d\sigma} &= 0\end{aligned}$$

So, given an $\Gamma^\mu{}_{\alpha\beta}$ and $g_{\alpha\beta}$, we can work out the equations in a given reference frame.

It turns out that we can simplify the calculation even further: $\Gamma^\mu{}_{\alpha\beta}$ can be found from $g_{\alpha\beta}$. Taking the partial derivative of the metric, we have that

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda}=\frac{\partial^2 y^\alpha}{\partial x^\lambda\partial x^\mu}\frac{\partial y^\beta}{\partial x^\nu}\eta_{\alpha\beta}+\frac{\partial^2 y^\beta}{\partial x^\lambda\partial x^\nu}\frac{\partial y^\alpha}{\partial x^\mu}\eta_{\alpha\beta}$$

From the definitiion of $\Gamma^\mu{}_{\alpha\beta}$ we can replace it in the above expression

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda}=\Gamma^\gamma{}_{\lambda\mu}\frac{\partial y^\alpha}{\partial x^\gamma}\frac{\partial y^\beta}{\partial x^\nu}\eta_{\alpha\beta}+\Gamma^\gamma{}_{\lambda\nu}\frac{\partial y^\beta}{\partial x^\gamma}\frac{\partial y^\alpha}{\partial x^\mu}\eta_{\alpha\beta}$$

which can be written as

$$\frac{\partial g_{\mu\nu}}{\partial x^\lambda}=\Gamma^\gamma{}_{\lambda\mu}g_{\gamma\nu}+\Gamma^\gamma{}_{\lambda\nu}g_{\gamma\mu}$$

We can now permute indices and add them to solve and find

$$\Gamma^\mu{}_{\alpha\beta}=\frac{1}{2}g^{\mu\nu}\left(\frac{\partial g_{\alpha\nu}}{\partial x^\beta}+\frac{\partial g_{\nu\beta}}{\partial x^\alpha}-\frac{\partial g_{\alpha\beta}}{\partial x^\nu}\right)$$

Hence, as advertised, given a metric, $g_{\mu\nu}$, we can find the connection coefficents $\Gamma^\mu{}_{\alpha\beta}$ and then solve the geodesic equation.

It often useful to find the geodesic equations in terms of a variational principle. In fact, it is also a convenient method for, given a metric, calculating the connection coefficients. Consider a path in space time $x^\alpha(\lambda)$. We can define the proper time elapsed between two points on that curve, A and B, to be

$$c\tau_{AB}=\int_A^B Ld\lambda=\int_A^B d\lambda\sqrt{-\eta_{\alpha\beta}\dot{x}^\alpha\dot{x}^\beta}$$

The derivatives are taken with regards to λ. It is now possible to define an action for the path $x^\alpha(\lambda)$:

$$S=-mc^2\tau$$

and we can minimize this action to find the path which takes the most amount of proper time between points A and B. This path will be the geodesic. For example, if we choose $x^0=c\lambda=ct$ we have that

$$S=-mc\int dt\sqrt{c^2-\left(\frac{dx}{dt}\right)^2}$$

More generally, we can use the Euler Lagrange equations:

$$\frac{d}{d\lambda}\left(\frac{\partial L}{\partial \dot{x}^\alpha}\right) = \frac{\partial L}{\partial x^\alpha}$$

Given that L is independent of x^α (and $d\tau = Ld\lambda$) we have that the equation of motion becomes

$$\frac{\partial L}{\partial \dot{x}^\alpha} = -\frac{1}{2}\frac{1}{L}\eta_{\alpha\beta}\frac{dx^\beta}{d\lambda} = -\frac{1}{2}\eta_{\alpha\beta}\frac{dx^\beta}{d\tau}$$

The resulting equations then become

$$\frac{d^2x^\alpha}{d\tau^2} = 0$$

Although the above action is reparametrization invariant (i.e. we can change our definition of λ and the action is unaffected), the square root is difficult to work with. It is easier to work with

$$\tilde{S} = -m\int d\lambda \eta_{\alpha\beta}\frac{dx^\alpha}{d\lambda}\frac{dx^\beta}{d\lambda} = m\int d\lambda L^2$$

The Euler-Lagrange equation then becomes

$$\frac{\partial L^2}{\partial x^\mu} - \frac{d}{d\lambda}\left(\frac{\partial L^2}{\partial \dot{x}^\mu}\right) = -2\frac{dL}{d\lambda}\frac{\partial L}{\partial \dot{x}^\mu}$$

Again, if we choose the affine parameter λ to be linear in τ we have that the right hand side becomes

$$\frac{dL}{d\lambda} = \frac{d}{d\lambda}\left(c\frac{d\tau}{d\lambda}\right) = 0$$

So, for a massive particle it makes sense to choose $\lambda = \tau$. Finally we, can rederive the geodesic equation from the action principle as above, but in the presence of a gravitational field. We now have

$$S = m\int d\lambda g_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = m\int d\lambda L^2$$

and can apply the Euler-Lagrange equation to obtain the Geodesic equation in a general frame.

5 Coordinate Transformations and Metrics

We have seen that coordinate transformations to accelerated reference frames lead to non-trivial geodesic equations. We have considered one such transformation when deriving the gravitational redshift. Let us now look at coordinate transformations in more detail.

You have extensively studied coordinate transformations between inertial reference frames (in the absence of gravitational fields) in Special Relativity. For example, consider a coordinate transformation along the z-axis from a stationary frame to one moving at a velocity v. We have that the coordinate transformation can be expressed in a matrix form as

$$\begin{pmatrix} \tilde{x}^0 \\ \tilde{x}^1 \\ \tilde{x}^2 \\ \tilde{x}^3 \end{pmatrix} = \begin{pmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{pmatrix} \begin{pmatrix} x^0 \\ x^1 \\ x^2 \\ x^3 \end{pmatrix}$$

where $\beta = v/c$ and $\gamma = 1/\sqrt{1-\beta^2}$. This coordinate transformation is of the form

$$\tilde{x}^\mu = \tilde{x}^\mu(x^\nu)$$

and given that it is linear, we have that the Jacobian is simply:

$$\frac{\partial \tilde{x}^\mu}{\partial x^\nu} = \begin{pmatrix} \gamma & 0 & 0 & -\beta\gamma \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -\beta\gamma & 0 & 0 & \gamma \end{pmatrix}$$

Now, in special relativity we have that the space time interval

$$ds^2 = -c^2 dt^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2$$

is invariant under coordinate transformations. We can restate this as

$$ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta = \eta_{\alpha\beta} d\tilde{x}^\alpha d\tilde{x}^\beta$$

in other words, the transformation leaves the actual *form* of the space time interval invariant.

What happens if we now consider a coordinate transformation to an accelerating reference frame? Let us consider the simplest case, an accelerating reference frame, with acceleration g, along the x^3 direction. We have that the transformation between the old coordinates, (x^0, x^1, x^2, x^3) and the new coordinates, $(\tilde{x}^0, \tilde{x}^1, \tilde{x}^2, \tilde{x}^3)$, is given by

$$\begin{aligned} x^0 &= \frac{c^2}{g} e^{\frac{g\tilde{x}^3}{c^2}} \sinh(\frac{g}{c^2}\tilde{x}^0) \\ x^1 &= \tilde{x}^1 \\ x^2 &= \tilde{x}^2 \\ x^3 &= \frac{c^2}{g} e^{\frac{g\tilde{x}^3}{c^2}} \cosh(\frac{g}{c^2}\tilde{x}^0) \end{aligned}$$

We can transform the expression for the space-time interval to the new coordinate system:

$$\begin{aligned} ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta &= -(dx^0)^2 + (dx^1)^2 + (dx^2)^2 + (dx^3)^2 \\ &= -e^{2\frac{g\tilde{x}^3}{c^2}}(d\tilde{x}^0)^2 + (d\tilde{x}^1)^2 + (d\tilde{x}^2)^2 + e^{2\frac{g\tilde{x}^3}{c^2}}(d\tilde{x}^3)^2 \end{aligned}$$

Note that the equivalent gravitational potential has the form $\Phi = g\tilde{x}^3$ and so the interval takes the form

$$ds^2 = -e^{2\frac{\Phi}{c^2}}(d\tilde{x}^0)^2 + (d\tilde{x}^1)^2 + (d\tilde{x}^2)^2 + e^{2\frac{\Phi}{c^2}}(d\tilde{x}^3)^2 \tag{5}$$

In fact, this expression is valid in a more general setting than just a constant gravitational acceleration. A weak, static gravitational field, Φ is equivalent to a metric of this form.

For the remainder of this section let us familiarize ourselves a bit more with coordinate transformations and how they affect the metric. With a general coordinate transformation, $\tilde{x}^\alpha = \tilde{x}^\alpha(x^\beta)$ we find that the space time interval changes as

$$ds^2 = \eta_{\alpha\beta} dx^\alpha dx^\beta = \eta_{\alpha\beta} \frac{\partial x^\alpha}{\partial \tilde{x}^\mu} \frac{\partial x^\beta}{\partial \tilde{x}^\nu} d\tilde{x}^\mu d\tilde{x}^\nu \equiv g_{\mu\nu} d\tilde{x}^\mu d\tilde{x}^\nu$$

In other words, under a general coordinate transformation, we have that $\eta_{\mu\nu} \to g_{\mu\nu}$. We call the object $g_{\mu\nu}$ the metric.

The metric contains information about the geometry of space (and space-time) we are considering. It is instructive to work through a few examples. For example, consider a 2-D sheet plane, with coordinates (x, y). The interval on that plane is given by

$$ds^2 = dx^2 + dy^2$$

and hence the metric is very simple: it is a diagonal matrix with entries

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

We can transform to polar coordinates

$$\begin{aligned} x &= r \cos\theta \\ y &= r \sin\theta \end{aligned}$$

to find

$$ds^2 = dr^2 + r^2 d\theta^2$$

Note that now the metric is more complicated:

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & r^2 \end{pmatrix}$$

yet it *still describes a plane*. We could have considered a different surface, a sphere with unit radius. It is a two dimensional surface and hence needs two coordinates, (θ, ϕ). The infinitesimal interval is defined as

$$ds^2 = d\theta^2 + \sin^2(\theta) d\phi^2$$

with metric

$$g_{ij} = \begin{pmatrix} 1 & 0 \\ 0 & \sin^2(\theta) \end{pmatrix}$$

The geometry of the surface of a sphere is obviously very different to the geometry of a plane and it is in the metric that this information is encoded.

We are of course, interested in the geometry of 3-D space and of 4-D space time. So, for example, the interval (and metric) for Euclidean (or flat) 3-D space in Cartesian coordinates are

$$ds^2 = (dx^1)^2 + (dx^2)^2 + (dx^3)^2$$

and

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

In spherical coordinates

$$\begin{aligned} x^1 &= r\sin(\theta)\cos(\phi) \\ x^2 &= r\sin(\theta)\sin(\phi) \\ x^3 &= r\cos(\theta) \end{aligned}$$

we have that the interval (and metric) are

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2\sin^2(\theta)d\phi^2$$

and

$$g_{ij} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & r^2 & 0 \\ 0 & 0 & r^2\sin^2(\theta) \end{pmatrix}$$

Again, these two metrics (in Cartesian and spherical coordinates) describe *exactly* the same space.

We have already seen examples of space-time metrics above. The Minkowski metric and the metric of an accelerated observer (known as a Rindler metric).

Let us now consider two important metric. The first one is that of a Euclidean, homogeneous and isotropic spacetime. We have that

$$ds^2 = -c^2dt^2 + a^2(t)\left[(dx^1)^2 + (dx^2)^2 + (dx^3)^2\right] \tag{6}$$

which has a metric

$$g_{\mu\nu} = \begin{pmatrix} -1 & 0 & 0 & 0 \\ 0 & a^2(t) & 0 & 0 \\ 0 & 0 & a^2(t) & 0 \\ 0 & 0 & 0 & a^2(t) \end{pmatrix}$$

A particularly important metric is a static (i.e. time independent), spherically symmetric *Schwarzschild metric*:

$$ds^2 = -\left(1 - \frac{2GM}{c^2r}\right)c^2dt^2 + \left(1 - \frac{2GM}{c^2r}\right)^{-1}dr^2 + r^2d\theta^2 + r^2\sin^2(\theta)d\phi^2 \tag{7}$$

which has a metric

$$g_{\mu\nu} = \begin{pmatrix} -\left(1-\frac{2GM}{c^2 r}\right) & 0 & 0 & 0 \\ 0 & \left(1-\frac{2GM}{c^2 r}\right)^{-1} & 0 & 0 \\ 0 & 0 & r^2 & 0 \\ 0 & 0 & 0 & r^2\sin^2(\theta) \end{pmatrix}$$

This metric is of particular importance. It corresponds to the space time of a point like mass and can be used to describe the space time around stars, planets and black holes.

To finish, let us calculate the geodesics for the homogeneous metric to find out what the connection coefficients are. The action is:

$$L^2 = -c^2\dot{t}^2 + a^2(t)\sum(\dot{x}^i)^2$$

where $\dot{f} = \frac{df}{d\lambda}$. The Euler-Lagrange equations are:

$$\begin{aligned} \ddot{x}^0 + \frac{a}{c}\frac{da}{dt}\sum(\dot{x}^i)^2 &= 0 \\ \ddot{x}^i + 2\frac{1}{ac}\frac{da}{dt}\dot{x}^0\dot{x}^i &= 0 \end{aligned}$$

We can now read off the connection coefficients (and be careful not to over count with the factor of 2 in the second expression):

$$\begin{aligned} \Gamma^0{}_{ij} &= \frac{1}{c}a\frac{da}{dt}\delta_{ij} \\ \Gamma^i{}_{0j} &= \frac{1}{ac}\frac{da}{dt}\delta^i_j \end{aligned}$$

6 The Newtonian Limit and the Gravitational Redshift Revisited

Interestingly enough, with what we have done we can already start relating the geodesic equation with the Newtonian regime of gravity. Let us look at the case where the gravitational field is extremely weak and stationary and particles are moving at non-relativistic speeds so $v \ll c$. Let us start with the geodesic equation:

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{\alpha\beta}\frac{dx^\alpha}{d\tau}\frac{dx^\beta}{d\tau} = 0$$

The non-relativistic approximation means that $dx^i/d\tau \ll d(ct)/d\tau$ so the geodesic equation simplifies to

$$\frac{d^2x^\mu}{d\tau^2} + \Gamma^\mu_{00}\left(\frac{dx^0}{d\tau}\right)^2 \simeq 0$$

If the gravitational field is stationary we have that $\partial g_{\mu\nu}/\partial t = 0$ and we have

$$\Gamma^\mu_{00} = -\frac{1}{2}g^{\mu\lambda}\frac{\partial g_{00}}{\partial x^\lambda}$$

Now consider a weak field

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$

where we assume that $|h_{\mu\nu}| \ll 1$. We can expand Γ^{μ}_{00} to first order in $h_{\mu\nu}$ to find

$$\Gamma^{\mu}_{00} = -\frac{1}{2}\eta^{\mu\nu}\frac{\partial h_{00}}{\partial x^{\nu}}$$

Only the spatial parts of $\eta_{\mu\nu}$ survive (which are 1). Hence we have

$$\Gamma^{\mu}_{00} = -\frac{1}{2}\frac{\partial h_{00}}{\partial x^{\mu}}$$

The geodesic equation then becomes

$$\frac{d^2x^{\mu}}{d\tau^2} = \frac{1}{2}\left(\frac{dx^0}{d\tau}\right)^2 \frac{\partial h_{00}}{\partial x^{\nu}}$$

which in vectorial notation becomes

$$\frac{d^2\mathbf{r}}{d\tau^2} = \frac{1}{2}\left(\frac{dx^0}{d\tau}\right)^2 \nabla h_{00}$$

For small speeds $dt/d\tau \simeq 1$ and comparing with the Newtonian result $\frac{d^2\mathbf{r}}{dt^2} = -\nabla\Phi$ we see that

$$h_{00} = -\frac{2\Phi}{c^2}$$

Hence in the weak-field limit

$$g_{00} = -\left(1 + \frac{2\Phi}{c^2}\right)$$

Note that, if we take the metric that we found in equation 5 and expand to linear order (i.e. take the weak field limit), we get exactly the same expression.

We can rederive our expression for the gravitational redshift *directly* from the metric. Consider the proper time again

$$-ds^2 = c^2 d\tau^2 = -g_{\mu\nu}dx^{\mu}dx^{\nu}$$

and pick a stationary system so $dx^i = 0$. We then have

$$d\tau = \sqrt{-g_{00}}dt.$$

In the weak field case we have

$$d\tau \simeq \left(1 + \frac{\Phi}{c^2}\right) dt$$

Note that t and τ only coincide if $\Phi = 0$ so clocks run slowly in potential wells. We can now compare the rate of change at two points, A and B, to get

$$\frac{d\tau_A}{d\tau_B} = \sqrt{\frac{g_{00}(A)}{g_{00}(B)}}$$

which in the weak field limit gives the ratio of frequencies

$$\frac{\nu_B}{\nu_A} = \frac{d\tau_A}{d\tau_B} \simeq 1 - \frac{\Phi_B - \Phi_A}{c^2}$$

which is equivalent to the expression we found in Section 3. We can define the gravitational redshift

$$z_{grav} \equiv \frac{\nu_A - \nu_B}{\nu_A} = \frac{\Phi_B - \Phi_A}{c^2}$$

This is a very small effect- for the Sun it is $\sim 10^{-6}$.

One way to look for this effect is by studying the spectral lines emitted from atoms which are very close to a massive body and hence deep into a gravitational potential. This effect has been observed in the Sun and white dwarfs but these observations are not very accurate. A classic test of the gravitational redshift was undertaken by Pound and Rebka in 1960 at Harvard. They used a 22.5 metre high tower where they placed an unstable nucleus, Fe^{57} at the top and the bottom. The nucleus (at the top) would emit gamma-rays with a certain frequency related to their energy. These rays would fall to the bottom and interact with the Fe^{57} there. If the gamma rays of the observer were the same as the emitter, the Fe^{57} at the bottom of the shaft would react. But because of the gravitational redshift, the frequency was shifted and the absorption was less efficient. By changing the velocity of the source at the top of the tower, the experimenters could compensate for the gravitational effect and measure it to within 1%.

Better measurements of the gravitational redshifting of light can be obtained on (or near) the Earth where, even though the gravitational field is much, much weaker, there is the possibility to make very precise measurements. One way to do this is to send a rocket up into orbit with a hydrogen-maser clock and emitting pulses to a ground station. At an altitude of 10^4 Km, the change in gravitational potential will be $gh/c^2 \simeq 10^{-10}$. Note that this effect is minute, almost 5 orders of magnitude smaller than the simple Doppler effect due to the motion of the rocket. Yet it is still possible to constrain the effect to within 0.002%.

7 Orbits I: the Perihelion of Mercury

It is now time to revisit the two body problem. We have already worked this out for the Newtonian case but we can now see what happens if we consider the more general case. Strictly speaking we will be studying the motion of mass μ in a central potential sourced by a mass M.

We can use the Schwarzschild metric that we introduced in equation 7 of the previous section. The total mass is M and the reduced mass is μ. The action for the geodesic equation for $(t(\lambda), r(\lambda), \theta(\lambda), \phi(\lambda))$ in this metric is

$$L^2 = \left(1 - \frac{2GM}{rc^2}\right) c^2 \dot{t}^2 - \frac{\dot{r}^2}{1 - \frac{2GM}{rc^2}} - r^2(\dot{\theta}^2 + \sin^2\theta\dot{\phi}^2)$$

The angular Euler-Lagrange equations are exactly as in the Newtonian two body problem we previously solved

$$\begin{aligned}\frac{d}{d\lambda}(2r^2\dot{\theta}) &= 2r^2\sin\theta\cos\theta\dot{\phi}^2\\ \frac{d}{d\lambda}(2r^2\sin^2\theta\dot{\phi}) &= 0\end{aligned}$$

and we can solve them in the same way, placing the orbit on the equatorial plane, choosing integration constants such that $\dot{\theta} = 0$ so that

$$r^2\dot{\phi} = \frac{J}{\mu}$$

The timelike component of the geodesic obeys

$$\frac{d}{d\lambda}\left[c^2\left(1 - \frac{2GM}{rc^2}\right)\dot{t}\right] = 0$$

which can be integrated to give

$$\left(1 - \frac{2GM}{rc^2}\right)\dot{t} = k$$

For a massive particle we have $L^2 = c^2$

$$c^2 = \frac{c^2k^2 - \dot{r}^2}{\left(1 - \frac{2GM}{c^2r}\right)} - \frac{J^2}{\mu^2r^2}$$

which can be rewritten

$$\dot{r}^2 + \left(1 - \frac{2GM}{c^2r}\right)\frac{J^2}{\mu^2r^2} = c^2k^2 - c^2 + \frac{2GM}{r}$$

Rearranging we find

$$\dot{r}^2 + \frac{J^2}{\mu^2r^2} - \frac{2GM}{r} - \frac{2GMJ^2}{\mu^2r^3c^2} = \text{constant} \tag{8}$$

We can compare this expression with the one we found in the Newtonian case in equation 2:

$$\dot{r}^2 + \frac{h^2}{r^2} - \frac{2GM}{r} = \text{constant}$$

There is an extra term in the General Relativistic case. Furthermore, in the Newtonian case we are taking derivatives with regards to t while in the General Relativistic case we are using the affine parameter λ.

We would now like to find the orbits of motion in this system. As in the Newtonian case, using conservation of angular momentum we can change the independent variable from λ to ϕ. Furthermore, we can transform to $u = 1/r$ so that

$$\dot{r} = \frac{dr}{d\phi}\frac{d\phi}{d\lambda} = -\frac{1}{u^2}\frac{du}{d\phi}\frac{J}{\mu}u^2 = -\frac{J}{\mu}\frac{du}{d\phi}$$

Dividing through by J^2/μ^2, equation 8 becomes

$$\left(\frac{du}{d\phi}\right)^2 + u^2 - \frac{2GM\mu^2}{J^2}u - \frac{2GM}{c^2}u^3 = constant$$

Differentiating by ϕ and dividing by $2\frac{du}{d\phi}$ we find

$$\frac{d^2u}{d\phi^2} + u = \frac{Gm\mu^2}{J^2} + \frac{3GM}{c^2}u^2$$

You will see that this equation has a very similar form to equation 3 with an extra bit. It is useful to rescale u so as to assess how important the correction is. Define $U = \frac{J^2}{GM\mu^2}u$. We then have

$$\frac{d^2U}{d\phi^2} + U = 1 + \epsilon U^2 \tag{9}$$

with $\epsilon \equiv 3G^2M^2\mu^2/J^2c^2$. We will see that ϵ in the case of Mercury, is of order 10^{-7}, so very small. We therefore assume that U can be split into a Newtonian part, U_0 and a small, general relativistic correction, U_1. We have that $U_0 = 1 + e\cos(\phi)$ which we can plug into equation 9 to find (to lowest order):

$$\frac{d^2U_1}{d\phi^2} + U_1 = \epsilon U_0^2 = \epsilon[1 + 2e\cos(\phi) + e^2\cos^2(\phi)] = \epsilon[1 + \frac{e^2}{2} + 2e\cos(\phi) + \frac{e^2}{2}\cos(2\phi)]$$

The complementary function is as before but the particular integral takes the form

$$U_1 = \epsilon\left[\left(1 + \frac{e^2}{2}\right) + e\phi\sin(\phi) - \frac{e^2}{6}\cos(2\phi)\right]$$

With time, the dominant term will be proportional to ϕ so that, adding the complementary function and the particular integral we have

$$U \simeq 1 + e\cos(\phi) + \epsilon e\phi\sin(\phi)$$

This corresponds to the Taylor expansion of

$$U \simeq 1 + e\cos[\phi(1-\epsilon)]$$

I.e. the period of the orbit is now $2\pi/(1-\epsilon)$ and *not* 2π. The orbit does not close in on itself. We can work out what this correction is for Mercury. Taking $M \simeq 2\times10^{30}$ kg, the orbital period $T = 88$ days and the mean orbital radius $r = 5.8\times10^{10}$ m, we find $\epsilon \simeq 10^{-7}$ so that precession rate is approximately 43 arc seconds per century. This effect, first detected by Le Verrier in the mid 19th century is obscured by a number of other effects. The precession of the equinoxes of the coordinate systems contributes to about 5025" per century while the other planets contribute about 531" per century. The Sun also has a quadropole moment which contributes a further 0.025" per century. Taking all these effects into account still leaves a precession of $\triangle\theta$ for which the current best estimate is

$$\triangle\theta = 42.969'' \pm 0.0052'' \text{ per century}$$

The prediction from General Relativity is

$$\triangle\theta \simeq 42.98'' \text{ per century}$$

8 Orbits II: Gravitational Lensing and the Shapiro Time Delay

We now want to study what happens to a light ray propagating in a gravitational field. We first work out what happens in a Newtonian universe; we are going to model a photon as a massive particle travelling at the speed of light, c with angular momentum per unit mass $h = cR$. Recall that $u = 1/r$ satisfies

$$\frac{d^2u}{d\phi^2} + u = \frac{GM}{h^2} = \frac{GM}{c^2R^2}$$

We have solved it before for closed orbits but we now want to pick integration constants that lead to unbounded orbits:

$$u = \frac{\sin\phi}{R} + \frac{GM}{c^2R^2}$$

where R is the distance of closest approach in the absence of gravity. Consider the asymptotic behaviour: when $r \to \infty$ we have $u \to 0$ which gives us two solutions for ϕ: $\phi_- = -GM/(c^2R)$ and $\phi_+ = \pi + GM/(c^2R)$. The total deflection is

$$\Delta\phi_N = \frac{2GM}{c^2R}$$

We can now repeat the calculation for in the relativistic case. As in Section 7, we have that the geodesic equation for a light ray must be parametrized in terms of an affine parameter and *not* proper time. We can use some of the results found in 7 (and, once again, taking $\theta = \pi/2$)

$$\begin{aligned} r^2\dot{\phi} &= h \\ \left(1 - \frac{2GM}{rc^2}\right)\dot{t} &= k \end{aligned}$$

For a massless particle we have $L^2 = 0$

$$0 = \left(1 - \frac{2GM}{c^2r}\right)c^2\dot{t}^2 - \frac{\dot{r}^2}{\left(1 - \frac{2GM}{c^2r}\right)} - \frac{h^2}{r^2}$$

which can be rewritten in terms of u as

$$h^2\left(\frac{du}{d\phi}\right)^2 = c^2k^2 - h^2u^2 + \frac{2GM}{c^2}h^2u^3$$

which, when differentiated gives

$$\frac{d^2u}{d\phi^2} + u = \frac{3GM}{c^2}u^2$$

Again, we can treat the right hand side as a small perturbation to the orbit

$$u_0 = \frac{\sin\phi}{R} \tag{10}$$

This equation corresponds to a straight line. The first order equation is

$$\frac{d^2u_1}{d\phi^2} + u_1 = \frac{3GM}{c^2R^2}\sin^2\phi = \frac{3GM}{2c^2R^2}(1 - \cos 2\phi)$$

which combined with u_0 leads to the complete, first order, solution

$$u = \frac{\sin\phi}{R} + \frac{3GM}{2c^2R^2}\left(1 + \frac{1}{3}\cos 2\phi\right)$$

At large distances, $u \to 0$ and assuming $\sin\phi \simeq \phi$ we have two possible solutions for ϕ: $\phi_- = -\frac{2GM}{c^2R}$ and $\phi_+ = \pi + \frac{2GM}{c^2R}$. The total deflection is then

$$\Delta\phi_{GR} = \frac{4GM}{c^2R}$$

We find that $\Delta\phi_{GR} = 2\Delta\phi_N$.

For a light ray grazing the limb of the Sun, the deflection will be $\theta \simeq 1.75$", famously measured by Arthur Eddington during his Eclipse expedition in 1919. The tightest observational constraint come from observations due to Shapiro, David, Lebach and Gregory who used around 2500 days worth of observation taken over 20 years- they used 87 VLBI sites and 541 radio sources yielding more than 1.7×10^6 observations and and obtained a constraint on θ:

$$\theta = (0.99992 \pm 0.00023) \times 1.75''$$

which is 3 orders of magnitude better than Eddington's original observations.

Another relativistic effect involving light rays is the *Shapiro time delay*, first proposed in 1964. Again, take the Schwarzschild metric in the equatorial plane and apply to electromagnetic wave (radar) propagating at the speed of light. We have that $ds^2 = 0$ so

$$0 = \left(1 - \frac{2GM}{c^2r}\right)c^2dt^2 - \frac{dr^2}{\left(1 - \frac{2GM}{c^2r}\right)} - r^2d\phi^2$$

Take the unperturbed solution, given in equation 10. We have

$$-\frac{dr}{r^2} = \frac{\cos(\phi)}{R}d\phi$$

which can be used to find

$$r^2d\phi^2 = dr^2\tan^2\phi = \frac{R^2dr^2}{r^2 - R^2}$$

The metric can then be rewritten as

$$c^2dt^2 = dr^2\left[\left(1 - \frac{2GM}{c^2r}\right)^{-2} + \left(1 - \frac{2GM}{c^2r}\right)^{-1}\frac{R^2}{r^2 - R^2}\right]$$

We can now expand to first order in $GM/(rc^2)$ to find

$$cdt = \pm\frac{rdr}{\sqrt{r^2 - R^2}}\left[1 + \frac{2GM}{rc^2} - \frac{GMR^2}{r^3c^2}\right]$$

This expression can be easily integrate between points A and B to give

$$c\Delta t = \pm \left[\sqrt{r^2 - R^2} + \frac{2GM}{c^2} \ln \left(\sqrt{\frac{r^2}{R^2} - 1} + \frac{r}{R} \right) - \frac{GM}{c^2} \sqrt{1 - \frac{R^2}{r^2}} \right]_A^B \tag{11}$$

We can now apply this to a planetary system. Let us take A to be the Earth and B to be Venus. The expression in equation 11 for Δt is the coordinate elapsed, *not* the time elapsed on the Earth, $\Delta\tau$. To relate these two time intervals recall that

$$ds^2 = -c^2 d\tau^2$$

where $d\tau$ is the proper time elapsed at on the Earth. We can assume a circular orbit so that $dr = d\theta = 0$ but clearly we have $d\phi \neq 0$. We are then left with

$$c^2 d\tau^2 = \left(1 - \frac{2GM}{c^2 r_E}\right) c^2 dt^2 - r^2 d\phi^2$$

where r_E is the Earth-Sun distance. We can simplify to

$$d\tau = \sqrt{1 - \frac{2GM}{c^2 r_E} - \frac{r^2}{c^2}\left(\frac{d\phi}{dt}\right)^2}\, dt$$

On a circular orbit we can use Kepler's law

$$\left(\frac{d\phi}{dt}\right)^2 = \frac{GM}{r^3}$$

so we find

$$\Delta\tau = \sqrt{1 - \frac{2GM}{c^2 r_E} - \frac{GM}{c^2 r_E}}\Delta t = \sqrt{1 - \frac{3GM}{c^2 r_E}}\Delta t \simeq \left(1 - \frac{3GM}{2c^2 r_E}\right)\Delta t$$

To understand what one should expect, take $r \gg R$ and look at the expression to see which term dominates as $R \to 0$. The logarithmic term will diverge so that, when a light ray passes close to the source, there is a large time delay. So, by monitoring a regular pulse of light as it passes behind a massive body, one should see a large include in the period, a characteristic spike during the transit. Shapiro proposed an experiment where one would send light rays (or radar signals) from the Earth which would then be reflected off Venus and back. If the Earth and Venus are aligned, the Sun induces an effect on the order of microseconds. In fact, when receiving signals from distant satellites such Voyager and Pioneer, one *has to* include the Shapiro time delay effect in processing their signals. The best constraints are due to Bertotti, Iess and Tortora using radio links with Cassini in 2002 which give us

$$\triangle t = (1.00001 \pm 0.00001)\triangle t_{GR}$$

where $\triangle t_{GR}$ is the prediction from General Relativity.

9 The Equivalence Principle and General Covariance

The Equivalence Principle has led us to move away from preferred frames or even preferred coordinate systems. A modern theory of gravity must take that into account, i.e. it should be possible to write the laws of physics in a form which is true in any coordinate system. This is known as the *Principle of General Covariance.*

To implement General Covariance we have to learn a little bit more about geometry. For a start let us recall how we transform between different coordinate systems. Consider a coordinate transformation $\tilde{x}^\mu = \tilde{x}^\mu(x^\nu)$. The Jacobian matrix of the transformation is defined to be

$$\frac{\partial \tilde{x}^\mu}{\partial x^\nu}$$

Given another coordinate transformation $\bar{x}^\mu = \bar{x}^\mu(\tilde{x}^\nu)$, we can apply the *chain rule* to get:

$$\frac{\partial \bar{x}^\mu}{\partial x^\nu} = \frac{\partial \bar{x}^\mu}{\partial \tilde{x}^\alpha}\frac{\partial \tilde{x}^\alpha}{\partial x^\nu}$$

and

$$\frac{\partial x^\mu}{\partial \tilde{x}^\alpha}\frac{\partial \tilde{x}^\alpha}{\partial x^\nu} = \delta^\mu{}_\nu$$

How do different types of functions of x^μ transform under coordinate transformations. The simplest case is a *scalar field*, $\phi(x^\mu)$- it remains unchanged under a coordinate transformation. A simple example of a scalar is $d\tau$, which we used in Section 4 to construct the invariant action for the geodesic.

The next type of functions are *vectors fields.* Consider a curve in space time, parametrized by λ so $x^\alpha = x^\alpha(\lambda)$. The *tangent vector field* is given by

$$T^\alpha = \frac{dx^\alpha}{d\lambda}$$

Suppose we now change coordinates to $\tilde{x}^\alpha$. We now have that the tangent vector in these new coordinates is

$$\tilde{T}^\alpha = \frac{d\tilde{x}^\alpha}{d\lambda}. \tag{12}$$

Using the chain rule we have

$$\frac{d\tilde{x}^\alpha}{d\lambda} = \frac{\partial \tilde{x}^\alpha}{\partial x^\beta}\frac{dx^\beta}{d\lambda}$$

So the tangent vector field transforms as

$$\tilde{T}^\alpha = \frac{\partial \tilde{x}^\alpha}{\partial x^\beta}T^\beta$$

A vector field with the indices "up" (and which therefore transforms in this way) is known as a *contravariant* vector field.

There is a different type of vector field, with an index "down" which is known as a *covariant* vector field. An example is

$$F_\alpha = \frac{\partial f}{\partial x^\alpha}$$

i.e. the gradient of a function $f(x)$. The chain rule gives

$$\frac{\partial f}{\partial x^\alpha} = \frac{\partial \tilde{x}^\beta}{\partial x^\alpha}\frac{\partial f}{\partial \tilde{x}^\beta}$$

So definining

$$\tilde{F}_\alpha = \frac{\partial f}{\partial \tilde{x}^\alpha}$$

we have

$$\tilde{F}_\alpha = \frac{\partial x^\beta}{\partial \tilde{x}^\alpha}F_\beta$$

Note how the transformation matrix is the inverse of the one for contravariant tensors. This of course, means that if you contract something with an "up" index with something with a "down" index you have

$$\tilde{T}^\alpha \tilde{F}_\alpha = \frac{\partial \tilde{x}^\alpha}{\partial x^\beta}\frac{\partial x^\gamma}{\partial \tilde{x}^\alpha}T^\beta F_\gamma = \delta^\gamma{}_\beta T^\beta F_\gamma = T^\beta F_\beta$$

i.e. the resulting object is a scalar and unchanged by a coordinate transformation.

It should be obvious that we can generalize this to objects with arbitrary number of "up" and "down" indices. These objects are known as *tensors*. We have already had to deal with one of them, the metric. The metric is a 2nd rank tensor and transforms as

$$\tilde{g}_{\alpha\beta} = \frac{\partial x^\mu}{\partial \tilde{x}^\alpha}\frac{\partial x^\nu}{\partial \tilde{x}^\beta}g_{\mu\nu}$$

We can also define the inverse of the metric which is simply the *contravariant* version of the metric, $g^{\alpha\beta}$ and satisfies:

$$g^{\alpha\mu}g_{\mu\beta} = \delta^\alpha{}_\beta$$

We can generalize to an arbitrary tensor. For example a 2nd rank contravariant tensor will be represented as $M_{\alpha\beta}$, a 2nd rank covariant tensor will be of the form $N^{\alpha\beta}$ and a 2nd rank mixed tensor will have the form $O^\alpha{}_\beta$. We can have higher order tensors (and we will come across one later on). For example a 4th order mixed rank tensor will be of the form $R^\alpha{}_{\beta\mu\nu}$.

We have now constructed this array of objects and wish to do operations on them. The most important operation we need to do is differentiation. Let us see why we can't use normal derivatives. We've already seen that $F_\alpha = \frac{\partial f}{\partial x^\alpha}$ transforms in the correct way. Let us now consider the 2nd derivative of F_α:

$$\frac{\partial F_\alpha}{\partial x^\beta}$$

Again let us consider a change of coordinates:

$$\begin{aligned}\frac{\partial F_\alpha}{\partial x^\beta} &= \frac{\partial^2 f}{\partial x^\alpha \partial x^\beta} = \frac{\partial \tilde{x}^\gamma}{\partial x^\alpha}\frac{\partial}{\partial \tilde{x}^\gamma}\left(\frac{\partial \tilde{x}^\delta}{\partial x^\beta}\frac{\partial f}{\partial \tilde{x}^\delta}\right) \\ &= \frac{\partial \tilde{x}^\gamma}{\partial x^\alpha}\frac{\partial \tilde{x}^\delta}{\partial x^\beta}\frac{\partial^2 f}{\partial \tilde{x}^\gamma \partial \tilde{x}^\delta} + \frac{\partial^2 \tilde{x}^\delta}{\partial x^\alpha \partial x^\beta}\frac{\partial f}{\partial \tilde{x}^\delta}\end{aligned}$$

If we were working with a covariant tensor we wouldn't have the extra term. So the normal, 2^{nd} derivative of a scalar (i.e. the Hessian) is not a 2^{nd} covariant tensor.

We can define the *covariant derivative*, ∇_α which obeys the following properties

- $\nabla_\alpha f = \frac{\partial f}{\partial x^\alpha}$
- It obeys the Liebnitz rule

  $$\nabla_\alpha(MN) = (\nabla_\alpha M)N + M(\nabla_\alpha N)$$

  for any two tensors M and N (we have hidden the indices).
- ∇_α commutes with contractions between indices.

We can construct such an operator as follows. It is a normal derivative when applied to a scalar. Applied to contravariant or covariant vector fields it acts as

$$\begin{aligned}\nabla_\alpha V^\beta &= \partial_\alpha V^\beta + \Gamma^\beta{}_{\alpha\mu}V^\mu \\ \nabla_\alpha U_\beta &= \partial_\alpha U_\beta - \Gamma^\mu{}_{\alpha\beta}U_\mu\end{aligned}$$

The objects that we have denoted by $\Gamma^\mu{}_{\alpha\beta}$ are *exactly* the connection coefficients that we came across when constructing the geodesic equations. We can use them to construct the covariant derivatives of 2^{nd} rank tensors too. So, for example we have

$$\begin{aligned}\nabla_\alpha M_{\mu\nu} &= \partial_\alpha M_{\mu\nu} - \Gamma^\beta{}_{\alpha\mu}M_{\beta\nu} - \Gamma^\beta{}_{\alpha\nu}M_{\mu\beta} \\ \nabla_\alpha N^{\mu\nu} &= \partial_\alpha N^{\mu\nu} + \Gamma^\mu{}_{\alpha\beta}N^{\beta\nu} + \Gamma^\nu{}_{\alpha\beta}N^{\mu\beta} \\ \nabla_\alpha O^\mu{}_\nu &= \partial_\alpha O^\mu{}_\nu + \Gamma^\mu{}_{\alpha\beta}O^\beta{}_\nu - \Gamma^\beta{}_{\alpha\nu}O^\mu{}_\beta\end{aligned}$$

Finally, the covariant derivative constructed in this way satisfies the *metricity* condition:

$$\nabla_\alpha g_{\mu\nu} = 0$$

Now let us revisit our curve on space-time $x^\alpha = x^\alpha(\lambda)$. We can define the absolute derivative of a vector V^μ along that path to be

$$\frac{DV^\mu}{D\lambda} \equiv T^\alpha \nabla_\alpha V^\mu$$

where the tangent vector T^α is defined in equation 12. We say that the vector V^μ is *parallely transported* along that path if

$$\frac{DV^\mu}{D\lambda} = 0$$

This is a differential equation for V^μ. We can start at a point $x^\alpha(\lambda = 0)$ and integrate to find the value V^μ at, for example, the point $x^\alpha(\lambda = 1)$. The result is *path dependent*. The parallel transport equation applied to T^α is

$$\frac{DT^\alpha}{D\lambda} = 0$$

can be rewritten as

$$\frac{d^2x^\alpha}{d\lambda^2} + \Gamma^\alpha{}_{\mu\nu}\frac{dx^\mu}{d\lambda}\frac{dx^\nu}{d\lambda} = 0$$

Note that this is nothing more than the geodesic equation we found in equation 4.

We now have the tools to construct laws of physics which are invariant under coordinate transformations. We need only apply the following two rules:

1. Wherever we see the Minkowski metric, $\eta_{\alpha\beta}$, replace by a general metric, $g_{\alpha\beta}$.
2. Wherever we see a partial derivative, $\frac{\partial}{\partial x^\alpha}$, replace by a covariant derivative, ∇_α.

Let us apply this prescription to Newton's 2$^{\text{nd}}$ law applied in special relativity

$$\frac{d(mV^\alpha)}{d\tau} = F^\alpha$$

If it is to be coordinate invariant, we have

$$\frac{D(mV^\alpha)}{D\tau} = F^\alpha$$

where the total derivative is defined in terms of the covariant derivative above.

10 The Curvature of Space-Time: Riemann Curvature Tensor

We have seen in previous lectures that by performing a coordinate transformation, it is possible to remove the effect of gravity locally. Such a set of coordinates is known as the *Local Inertial Frame* (LIF). They are, for example, the fixed coordinates defined relative to an object in free fall. But given that we can always transform to a LIF, how can we tell if we are in the presence of a gravitational field?

When we transform to a LIF, we find a coordinate system such that $g_{\alpha\beta} \to \eta_{\alpha\beta}$ and the connection coefficients, Γ (which are built of first derivatives of $g_{\alpha\beta}$) vanish. Hence, the gravitational field must arise through *second*-derivatives of $g_{\alpha\beta}$, i.e. neighbouring points will feel different accelerations because the connection coefficients differ. We can schematically think of $\Gamma(x) \sim \partial_\alpha g_{\mu\nu}(x)$ and if we Taylor expand around a point x we have

$$\Gamma(x + \Delta x) \simeq \partial_\alpha g_{\mu\nu}(x) + \partial_\beta\partial_\alpha g_{\mu\nu}(x)\Delta x^\beta$$

Hence the forces (which come into the geodesic equation via the connection coefficients) will be different if $\partial_\beta\partial_\alpha g_{\mu\nu}(x) \neq 0$. This is a 4$^{\text{th}}$ rank object although not necessarily a tensor (note

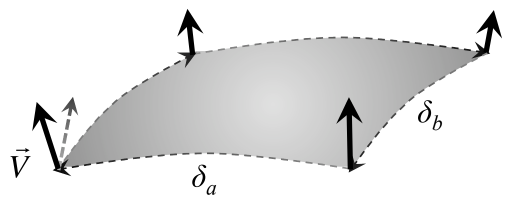

Figure 3: Parallel transport of vector V

that it is built out of normal derivatives, not covariant ones). We need something of this form which *is* a tensor to relate to the gravitational field and we can find it if we revisit our equation for parallel transport.

Recall that our equation for parallel transport is

$$\frac{DV^\mu}{D\lambda} = 0$$

which can be explicitely written as

$$\frac{dV^\mu}{d\lambda} = -\Gamma^\mu{}_{\alpha\beta}\frac{dx^\alpha}{d\lambda}V^\beta$$

We have then that the change around an infinitesimal length, δx^β is

$$\delta V^\mu = -\Gamma^\mu{}_{\alpha\beta}V^\alpha \delta x^\beta$$

Let us now parallely transport the vector V^μ around a parallelogram with sides δa^α and δb^β. We have that the total change is given by adding up the four contributions:

$$\begin{aligned}\delta V^\mu = &-\Gamma^\mu{}_{\alpha\beta}(x)V^\alpha(x)\delta a^\beta - \Gamma^\mu{}_{\alpha\beta}(x+\delta a)V^\alpha(x+\delta a)\delta b^\beta \\ &+\Gamma^\mu{}_{\alpha\beta}(x+\delta b)V^\alpha(x+\delta b)\delta a^\beta + \Gamma^\mu{}_{\alpha\beta}(x)V^\alpha(x)\delta b^\beta\end{aligned}$$

Note that we are not progressing sequentially around each corner of the square. We can now take the Taylor expansion of the middle two terms:

$$\delta V^\mu = -\frac{\partial(\Gamma^\mu{}_{\alpha\beta}V^\alpha)}{\partial x^\nu}\delta a^\nu \delta b^\beta + \frac{\partial(\Gamma^\mu{}_{\alpha\beta}V^\beta)}{\partial x^\nu}\delta a^\alpha \delta b^\nu$$

If we now relabel the 2$^{\text{nd}}$ term, $\alpha \leftrightarrow \nu$, then $\beta \leftrightarrow \alpha$ and using the product rule, we have:

$$\delta V^\mu = (\partial_\nu \Gamma^\mu{}_{\alpha\beta}V^\alpha + \Gamma^\mu{}_{\alpha\beta}\partial_\nu V^\alpha - \partial_\beta \Gamma^\mu{}_{\alpha\nu}V^\alpha - \Gamma^\mu{}_{\alpha\nu}\partial_\beta V^\alpha)\delta a^\nu \delta b^\beta$$

We can now use the parallel transport equation above to replace $\partial_\nu V^\alpha$ and $\partial_\beta V^\alpha$ and we find

$$\delta V^\mu = -R^\mu{}_{\nu\alpha\beta}V^\nu \delta a^\alpha \delta b^\beta$$

where

$$R^{\mu}{}_{\nu\alpha\beta} \equiv \partial_\alpha \Gamma^{\mu}{}_{\beta\nu} - \partial_\beta \Gamma^{\mu}{}_{\alpha\nu} + \Gamma^{\mu}{}_{\alpha\epsilon}\Gamma^{\epsilon}{}_{\nu\beta} - \Gamma^{\mu}{}_{\epsilon\beta}\Gamma^{\epsilon}{}_{\nu\alpha} \tag{13}$$

We have that $R^{\mu}{}_{\nu\alpha\beta}$ is the *Riemann curvature tensor* and it quantifies the curvature of a surface (in this case space-time). If there was no curvature, the parallel transport around a closed loop would bring a vector back onto itself. $R^{\mu}_{\nu\alpha\beta}$ is indeed a 4^{th} rank tensor and depends on the second derivative. It should therefore be useful for teasing out the effects of a gravitational field. In fact we can define the Riemann curvature tensor in terms of covariant derivatives and vectors through:

$$(\nabla_\alpha \nabla_\beta - \nabla_\beta \nabla_\alpha) V^\mu = R^{\mu}{}_{\nu\alpha\beta} V^\nu$$

The Riemann tensor satisfies a number of symmetry properties:

$$\begin{aligned} R^{\mu}{}_{\nu\alpha\beta} &= -R^{\mu}{}_{\nu\beta\alpha} \\ R_{\mu\nu\alpha\beta} &= -R_{\nu\mu\alpha\beta} \\ R_{\mu\nu\alpha\beta} &= R_{\alpha\beta\mu\nu} \\ R^{\mu}{}_{\nu\alpha\beta} + R^{\mu}{}_{\alpha\beta\nu} + R^{\mu}{}_{\beta\nu\alpha} &= 0 \end{aligned}$$

where the first index has been lowered using the metric: $R_{\mu\nu\alpha\beta} = g_{\mu\epsilon} R^{\epsilon}{}_{\nu\alpha\beta}$. We also have the *Bianchi* identity:

$$\nabla_\gamma R^{\mu}{}_{\nu\alpha\beta} + \nabla_\beta R^{\mu}{}_{\nu\gamma\alpha} + \nabla_\alpha R^{\mu}{}_{\nu\beta\gamma} = 0,$$

Finally, we can define the *Ricci* tensor and scalar:

$$\begin{aligned} R_{\alpha\beta} &\equiv R^{\mu}{}_{\alpha\mu\beta} \\ R &\equiv g^{\alpha\beta} R_{\alpha\beta} \end{aligned}$$

11 Building the Einstein Field Equations

We now want to progress to the equations that tell us how the gravitational field is sourced. We may find a hint of how to construct the field equation from Newtonian gravity. In Newtonian gravity we have the Poisson equation:

$$\nabla^2 \Phi = 4\pi G \rho$$

Consider now two neighbouring particles, x^i and $\tilde{x}^i = x^i + N^i$ in this gravitational field. Newton's 2^{nd} Law gives us

$$\begin{aligned} \frac{d^2 x^i}{dt^2} &= -\partial_i \Phi(\mathbf{x}) \\ \frac{d^2 \tilde{x}^i}{dt^2} &= -\partial_i \Phi(\mathbf{x} + \mathbf{N}) \end{aligned}$$

If we take the Taylor expansion of the 2nd equation we have, to lowest order:

$$\frac{d^2N^i}{dt^2} = -\partial_j\partial_i\Phi N^j$$

We can define the *tidal tensor*:

$$E_{ij} \equiv \partial_j\partial_i\Phi$$

so that

$$\frac{d^2N^i}{dt^2} + E_{ij}N^j = 0$$

(let us not worry about the fact that "up" and "down" indices don't match just this once). We can call this the *geodesic deviation equation* for Newtonian gravity. Now let us revisit the Poisson equation; we have that it can be rewritten as

$$E_{ii} = 4\pi G\rho$$

We can now use this link between the geodesic deviation equation and the Poisson equation to construct the appropriate field equations for General Relativity. In constructing geodesics, we saw that the metric, $g_{\alpha\beta}$ played the role of gravitational potentials and we now need a set of equations which are invariant under coordinate transformations. Consider now a family of geodesics $x^\alpha(\lambda, \sigma)$. We move *along* a geodesic by fixing σ and varying λ. We can move from one geodesic to the next one by fixing λ and varying σ. We have the vector which is tangent to a given geodesic is simply:

$$T^\alpha = \frac{dx^\alpha}{d\lambda}|_\sigma$$

while the vector which is orthogonal or normal to a geodesic at a point is

$$N^\alpha = \frac{dx^\alpha}{d\sigma}|_\lambda$$

We have that normal derivatives commute so $\partial_\sigma T^\alpha = \partial_\lambda N^\alpha$. We can think of λ as a time coordinate and σ as a spatial coordinate, which means we can pick a coordinate system: $(\lambda, \sigma, x^2, x^3)$. We then have that the tangent and normal vectors take a particularly simple form:

$$T^\alpha = \delta^\alpha_0 \qquad N^\alpha = \delta^\alpha_1$$

Again, from the commutation of the normal derivatives we have $N^\beta\partial_\beta T^\alpha - T^\beta\partial_\beta N^\alpha = 0$ which remains true if we replace the normal derivatives by covariant derivatives:

$$N^\beta\nabla_\beta T^\alpha - T^\beta\nabla_\beta N^\alpha = 0$$

Take the equation which relates the Riemann curvature tensor with the commutator of covariant derivatives:

$$(\nabla_\mu\nabla_\nu - \nabla_\nu\nabla_\mu)T^\alpha = R^\alpha{}_{\beta\mu\nu}T^\beta$$

Now contract it with N^α and T^β:

$$N^\mu T^\nu(\nabla_\mu\nabla_\nu T^\alpha - \nabla_\nu\nabla_\mu T^\alpha) = R^\alpha{}_{\beta\mu\nu}T^\beta T^\nu N^\mu \equiv E^\alpha{}_\mu N^\mu \tag{14}$$

We now have that

$$\frac{D^2N^\alpha}{D\lambda^2} = T^\mu\nabla_\mu(T^\nu\nabla_\nu N^\alpha)$$

which we can use the above commutation relation to rewrite as

$$\frac{D^2N^\alpha}{D\lambda^2} = T^\mu\nabla_\mu(N^\nu\nabla_\nu T^\alpha)$$

Using Leibnitz rule we have

$$\frac{D^2N^\alpha}{D\lambda^2} = T^\mu N^\nu\nabla_\mu\nabla_\nu T^\alpha + T^\mu\nabla_\mu N^\nu\nabla_\nu T^\alpha$$

If we add this equation to equation 14 (and use the geodesic equation for T^α) we find:

$$\begin{aligned}\frac{D^2N^\alpha}{D\lambda^2} + E^\alpha_\beta N^\beta &= T^\mu\nabla_\mu N^\beta\nabla_\beta T^\alpha + T^\mu N^\beta\nabla_\beta\nabla_\mu T^\alpha \\ &= T^\mu\nabla_\mu N^\beta\nabla_\beta T^\alpha + N^\beta\nabla_\beta(T^\mu\nabla_\mu T^\alpha) - N^\beta\nabla_\beta T^\mu\nabla_\mu T^\alpha \\ &= T^\mu\nabla_\mu N^\beta\nabla_\beta T^\alpha - T^\beta\nabla_\beta N^\mu\nabla_\mu T^\alpha = 0\end{aligned}$$

We have then that the geodesic deviation equation in general relativity has a similar form to that in Newtonian gravity but with a tidal tensor defined in equation 14. Clearly the Riemann curvature (or some reduced version of it) must play the role that $\nabla^2\Phi$ plays in the Newtonian gravity. In fact, in the Newtonian limit we have

$$E_{ii} = R^i{}_{0i0}$$

so the Newton Poisson equation is of the form

$$R_{\alpha\beta}T^\alpha T^\beta \simeq 4\pi G\rho$$

The form of the equation is schematically "Geometry $\simeq$ Energy", in other words, the matter will source the geometry in someway. This is giving us a hint that the general relativistic equation should be of the form

$$R_{\alpha\beta} \sim GT_{\alpha\beta}$$

where $T_{\alpha\beta}$ is a tensor which must be determined by the matter distribution.

12 The Energy-Momentum Tensor

We now need to construct the object, $T_{\alpha\beta}$ which will be used for the General Relativistic Field equations. Cleary it must involve ρ. From Special Relativity we know that, if we change to a moving frame, with velocity v and boost factor, γ, we have that $\rho \to \gamma^2\rho$. Which means that ρ transforms like a component of a 2^{nd} rank tensor and hence fits nicely in an object such as $T_{\alpha\beta}$. A first guess would be

$$T^{\alpha\beta} = \rho U^\alpha U^\beta$$

where U^α is the 4-velocity of the fluid, $U^\alpha \equiv \gamma(c, \mathbf{v})$. If we take the divergence of this tensor, we obtain a conservation equation:

$$\partial_\alpha T^{\alpha\beta} = 0$$

Setting $\gamma = 1$ we have two familiar conservation equations. First of all, conservation of mass:

$$\frac{\partial\rho}{\partial t} + \partial_i(\rho v^i) = 0$$

and momentum

$$\frac{\partial}{\partial t}(\rho v^i) + \partial_k(\rho v^i v^k) = 0$$

The latter equation can be reexpressed as the Euler equation for a pressureless fluid.

We want to represent a more general, perfect fluid, one that include pressure, P and has a form which is invariant under coordinate transformation- i.e. a proper tensor. This can be achieved with

$$T^{\alpha\beta} = (\rho + \frac{P}{c^2})U^\alpha U^\beta + Pg^{\alpha\beta}$$

where the 4-velocity of the fluid satisfies $U^\alpha U_\alpha = -c^2$. In the local rest frame of the fluid, we have $U = (c, \mathbf{0})$. The energy-momentum conservation equation now becomes

$$\nabla_\alpha T^{\alpha\beta} = 0$$

We can construct the energy-momentum tensor for just about anything: vector fields (like the electric and magnetic fields), gases of particles described by distribution functions, fermions, scalar fields, etc.

13 The Einstein equations and the Newtonian Limit

Let us now attempt to construct the field equations which are tensorial and which have at most 2^{nd} derivatives. The only tensors at our disposal are, $R^{\alpha\beta}$, $T^{\alpha\beta}$ and $g^{\alpha\beta}$. We also have two scalars $T \equiv g_{\alpha\beta}T^{\alpha\beta}$ and the Ricci scalar, R; it turns out that one of these will be redundant in what follows so we will discard T. The most general equation is

$$R^{\alpha\beta} = AT^{\alpha\beta} + Bg^{\alpha\beta} + CRg^{\alpha\beta} \tag{15}$$

where A, B and C are constants. From energy-momentum conservation we have that

$$\nabla_\alpha T^{\alpha\beta} = 0$$

and the metricity condition gives us

$$\nabla_\mu g^{\alpha\beta} = 0$$

If we take the Bianchi identity

$$\nabla_\gamma R^\mu{}_{\nu\alpha\beta} + \nabla_\beta R^\mu{}_{\nu\gamma\alpha} + \nabla_\alpha R^\mu{}_{\nu\beta\gamma} = 0,$$

contract μ with α and multiply by $g^{\nu\gamma}$ we find

$$2\nabla_\nu R^\nu{}_\beta - \nabla_\beta R = 0 \tag{16}$$

We can replace this expression in Equation 15 to find $C = \frac{1}{2}$. This allows us to define the *Einstein tensor*:

$$G^{\alpha\beta} \equiv R^{\alpha\beta} - \frac{1}{2} g^{\alpha\beta} R$$

We are left with two constants A and B so that

$$G^{\alpha\beta} = AT^{\alpha\beta} + Bg^{\alpha\beta}$$

What are these constants? Let us first focus on A. We can get an idea of where they come from by comparing the field equations with the Newton Poisson equation. To do so we have to make two approximations. First of all we need to consider the *weak field* limit of gravity so that we can expand the metric around a flat space, Minkowski space time:

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}$$

where $|h_{\mu\nu}| \ll 1$. Second, we will consider a time independent source with low speeds. The energy momentum tensor then becomes

$$T_{00} = \rho c^2 \quad T_{ij} \simeq 0$$

To first order in $h_{\mu\nu}$ the affine connections are

$$\Gamma^\alpha{}_{\mu\nu} \simeq \frac{1}{2} \eta^{\alpha\beta} (\partial_\mu h_{\beta\nu} + \partial_\nu h_{\mu\beta} - \partial_\beta h_{\mu\nu})$$

and the Riemann tensor becomes (note that we can discard products of Γs because they are 2^{nd} order)

$$R^\alpha{}_{\beta\mu\nu} = \frac{1}{2} \eta^{\alpha\gamma} (\partial_\beta \partial_\mu h_{\nu\gamma} - \partial_\gamma \partial_\mu h_{\nu\beta} - \partial_\beta \partial_\nu h_{\mu\gamma} + \partial_\nu \partial_\gamma h_{\mu\beta})$$

The Ricci tensor is then

$$R_{\beta\nu} = R^\alpha{}_{\beta\alpha\nu} = \frac{1}{2} \eta^{\alpha\gamma} (\partial_\beta \partial_\alpha h_{\nu\gamma} - \partial_\gamma \partial_\alpha h_{\nu\beta} - \partial_\beta \partial_\nu h_{\alpha\gamma} + \partial_\nu \partial_\gamma h_{\alpha\beta})$$

To find A we will focus on the G_{00} component of the field equations. This means we need R_{00}. We have chose a static source so that all derivatives with regards to 0 vanish. This leaves us with

$$R_{00} = -\frac{1}{2}\eta^{\alpha\gamma}\partial_\gamma\partial_\alpha h_{00} \simeq -\frac{1}{2}\nabla^2 h_{00}$$

Now taking $g_{00} \simeq -(1 + 2\frac{\Phi}{c^2})$ we have

$$R_{00} = \frac{1}{c^2}\nabla^2\Phi$$

For the Ricci scalar we can do the same. There is a trick we can use: assuming $|T_{ij}| \simeq 0$ (as declared above), and setting $B = 0$ we have $|G_{ij}| \simeq 0$ and so

$$R_{ij} \simeq \frac{1}{2}g_{ij}R = \frac{1}{2}\delta_{ij}R$$

The Ricci scalar is

$$R = g^{\alpha\beta}R_{\alpha\beta} \simeq \eta^{\alpha\beta}R_{\alpha\beta} = -R_{00} + R_{ii} = -R_{00} + \frac{3}{2}R. \tag{17}$$

So $2R_{00} = R$ and $G_{00} = R_{00} - \frac{1}{2}g_{00}R \simeq R_{00} + R_{00} = 2R_{00}$ We can now replace it all in the field equations:

$$\frac{2}{c^2}\nabla^2\Phi = A\rho c^2$$

If this is to agree with the Poisson equation we must have

$$A = \frac{8\pi G}{c^4}$$

Finally, it is a convention that we have $B = -\Lambda$ and we call Λ the *cosmological constant*. We then have that the Einstein field equations are:

$$G^{\alpha\beta} = \frac{8\pi G}{c^4}T^{\alpha\beta} - \Lambda g^{\alpha\beta} \tag{18}$$

We have completed our quest for a new theory of gravity. The set of *Einstein Field Equations* given in equations 18 and the geodesic equations, given in equations 4 replace Newton's Universal law of gravitation and Newton's 2$^{\text{nd}}$ law. In Einstein's theory, the picture is different- as the American physicist John Archibald Wheeler said: "Space tells matter how to move and matter tells space how to curve."

14 Black Holes

We now know how to, given a distribution of mass, derive the corresponding self-consistent metric. The Einstein Field Equations are, of course, a tangled mess of non-linear equations with 10 unknown functions of space time. It helps to consider symmetric configurations. We now look at one such configuration, that of a spherically symmetric, static metric in vacuum. If we write such a metric in spherical polar coordinates we have

$$ds^2 = -c^2 f(r)dt^2 + g(r)dr^2 - e(r)dtdr + h(r)(d\theta^2 + \sin^2\theta d\phi^2)$$

With a judicious redefinition of r and t we can eliminate $e(r)$ and $h(r)$ and we can then work with

$$ds^2 = -c^2 f(r)dt^2 + g(r)dr^2 + r^2(d\theta^2 + \sin^2\theta d\phi^2)$$

We will solve the Einstein Field Equations in empty space, where $T_{\mu\nu} = 0$. The equations can be rewritten as $R_{\mu\nu} = 0$ and so the challenge is to construct the Ricci tensor for such a space time with

$$\begin{aligned} g_{00} &= -f(r) \\ g_{rr} &= g(r) \\ g_{\theta\theta} &= r^2 \\ g_{\phi\phi} &= r^2\sin^2\theta \end{aligned}$$

There are 9 non-zero connection coefficients and these are:

$$\begin{aligned}
\Gamma^r{}_{00} &= -\frac{1}{2}g^{rr}\partial_r g_{00} = \frac{1}{2}\frac{f'}{g} \\
\Gamma^0{}_{0r} &= \frac{1}{2}g^{00}\partial_r g_{00} = -\frac{1}{2}\frac{f'}{f} \\
\Gamma^r{}_{rr} &= \frac{1}{2}g^{rr}\partial_r g_{rr} = \frac{1}{2}\frac{g'}{g} \\
\Gamma^r{}_{\theta\theta} &= -\frac{1}{2}g^{rr}\partial_r g_{\theta\theta} = -\frac{r}{g} \\
\Gamma^r{}_{\phi\phi} &= -\frac{1}{2}g^{rr}\partial_r g_{\phi\phi} = -\frac{r\sin^2\theta\theta}{g} \\
\Gamma^\theta{}_{\theta r} &= \frac{1}{2}g^{\theta\theta}\partial_r g_{\theta\theta} = \frac{1}{r} \\
\Gamma^\theta{}_{\phi\phi} &= -\frac{1}{2}g^{\theta\theta}\partial_\theta g_{\phi\phi} = -\sin\theta\cos\theta \\
\Gamma^\phi{}_{\phi r} &= \frac{1}{2}g^{\phi\phi}\partial_r g_{\phi\phi} = \frac{1}{r} \\
\Gamma^\phi{}_{\phi\theta} &= -\frac{1}{2}g^{\phi\phi}\partial_\theta g_{\phi\phi} = \frac{\cos\theta}{\sin\theta}
\end{aligned}$$

We now need expressions for $R_{\mu\nu}$ which are

$$\begin{aligned}
R_{00} &= \frac{1}{2}\frac{f''}{g} + \frac{1}{4}\frac{f'f'}{fg} - \frac{1}{4}\frac{f'g'}{g^2} + \frac{1}{2r}\frac{f'}{g} \\
R_{rr} &= \frac{1}{2}\frac{f''}{f} + \frac{1}{4}\frac{f'f'}{f^2} - \frac{1}{4}\frac{f'g'}{fg} + \frac{1}{2r}\frac{g'}{g} \\
R_{\theta\theta} &= \frac{1}{g} - 1 + \frac{r}{2g}\left(\frac{f'}{f} - \frac{g'}{g}\right) \\
R_{\phi\phi} &= \sin^2\theta R_{\theta\theta}
\end{aligned}$$

There are no off-diagonal terms and there are only three equations. We can now take the first two to find

$$\frac{g}{f}R_{00} + R_{rr} = -\frac{1}{r}\left(\frac{f'}{f} + \frac{g'}{g}\right) = 0$$

This easily solved to give $fg = \alpha$ where α is a constant. Replacing g in the $R_{\theta\theta} = 0$ we find

$$f + rf' = \alpha$$

which integrated gives us

$$f = \alpha(1 + \frac{\kappa}{r})$$

Matching this expression to the weak field limit (i.e. the Newtonian regime) we find $\alpha = 1$ and $\kappa = -2GM/c^2 r$ so that the final solution is the Schwarzschild metric:

$$ds^2 = -\left(1 - \frac{2GM}{c^2 r}\right)c^2 dt^2 + \left(1 - \frac{2GM}{c^2 r}\right)^{-1} dr^2 + r^2 d\theta^2 + r^2 \sin^2(\theta) d\phi^2$$

which we have used extensively throughout these lecture notes.

We have focused on the weak field region of this space time but let us try an understand a bit more about its peculiarities. For a start, something odd seems to happen at $r_S = 2GM/c^2$, known as the Schwarzschild radius: the metric seems to blow up. Nevertheless we know that the Ricci tensor is 0 and if we calculate the Riemann tensor we find that

$$R^{\alpha\beta\mu\nu}R_{\alpha\beta\mu\nu} = 12\frac{r_S^2}{r^4}$$

i.e. it is also finite. It turns out that at r_S we don't have a genuine space time singularity but a coordinate singularity. This doesn't mean that odd things don't happen at that, or near that point.

Let us consider geodesics in this space time. Using the geodesics that we derived above, we have that the radial equation is

$$\ddot{r} = -\frac{GM}{r^2} + \frac{h^2}{r^3}\left(1 - \frac{3GM}{c^2 r}\right)$$

We can find the stable minima of the left hand side (to which we can associate circular orbits) to find that they are at

$$r = \frac{2h^2 \pm \sqrt{4h^4 - 12r_S^2 c^2 h^2}}{2r_S c^2}$$

There is clearly a limit of h below which there is no solution and it is given by $h^2 = 3r_S^2 c^2$. This corresponds to the *Innermost Stable Circular Orbit*: $r_{ISCO} = 3r_s$. There are *no* circular orbits with smaller radii, all orbits are inspiralling towards r_S. Interestingly enough, outside this orbit, circular orbits *do* satisfy Kepler's law: $(d\phi/dt)^2 = GM/r^3$.

Let us now study what happens to an infalling particle. We can set $h = 0$ to find

$$\ddot{r} = -\frac{r_S c^2}{2r^2}$$

Taking $\dot{r} = 0$ at $r \to \infty$ we find an integral of motion

$$\dot{r}^2 = \frac{r_S c^2}{2r}$$

Integrating this equation, taking $\tau = 0$ at $r = r_0$ and defining $\xi = r/r_S$ we find

$$\frac{c\tau}{r_S} = \frac{2}{3}(\xi_0^{3/2} - \xi^{3/2})$$

Taking $xi = 0$ we find that it take a *finite* amount of proper time for a particle to reach $r = 0$ from any radius (within or without the Schwarazschild radius). If, however, we wish to find the amount of time elapsed for an observer at infinity, we need to integrate

$$\frac{dr}{dt} = -\left(\frac{r_s c^2}{r}\right)^{1/2}\left(1 - \frac{r_S}{r}\right)$$

which gives

$$\frac{ct}{r_S} = \frac{2}{3}(\xi_0^{3/2} - \xi^{3/2}) + 2(\xi_0^{1/2} - \xi^{1/2}) + \ln\left|\frac{(\xi^{1/2}+1)(\xi_0^{1/2}-1)}{(\xi^{1/2}-1)(\xi_0^{1/2}+1)}\right|$$

If we the endpoint to be the Schwarzschild radius (i.e. $\xi = 1$), we have that $ct \to \infty$. In other words, from an external observer it takes an infinite amount of time for the particle to fall in.

We can solve the geodesic equations for radial light rays by looking directly at the metric. We then have

$$\left(1 - \frac{r_S}{r}\right)^{1/2} cdt = \pm\left(1 - \frac{r_S}{r}\right)^{-1/2} dr$$

which can be rewritten as

$$c\frac{dt}{dr} = \pm(1 - \frac{r_S}{r})^{-1}$$

and integrated to give

$$ct = \pm r \pm r_s \ln|\frac{r}{r_S} - 1| + r_0$$

where r_0 is a constant of integration. With r_S we have the usual light cones with which we are familiar in Minkowski space. This is also approximately true for $r/r_S \gg 1$. But for r/r_S, the light cone "tips over" so that all forward moving particles necessarily move inwards towards $r = 0$. It is not possible to causally exit the Schwarzschild radius. The Schwarzschild radius works as a horizon beyond which we can't see anything. It is an *event* horizon.

Finally, such strong effects may still be at play even if the object is not a black hole. Very dense objects like neutrons stars can be a good laboratory for testing gravity. And indeed, *Binary pulsars* are incredibly useful astronomical objects that can be used to place very tight constraints on General Relativity. *Pulsars* are rapidly rotating neutrons stars that emit a beam of electromagnetic radiation, and were first observed in 1967. When these beams pass over the Earth, as the star rotates, we detect regular pulses of radiation. The first pulsar observed in a binary system was PSR B1913+16 in 1974, by Hulse and Taylor. The rotational period of this pulsar is about 59ms as it orbits around another neutron star. Binary pulsars are highly relativistic. For example the Hulse-Taylor pulsar precesses relativistically more than 30 000 times faster than the Mercury-Sun system. They are also a source of gravitational radiation, i.e. waves in space time that propagate away from the system and take energy away. We can predict how much energy in gravitational radiation a binary system will emit, in the context of General Relativity- it agrees almost perfectly with the angular momentum decay observed in the Hulse-Taylor pulsar. Indeed the Hulse-Taylor pulsar is an incredibly rich laboratory for General Relativity. At least 5 General Relativistic effects have been measured: orbital precession (also known as periastron advance), the rate of change of the orbital period, the gravitational redshift and two versions of the Shapiro time-delay effects.

15 Homogeneous and Isotropic Space- Times

The Einstein Field Equations are a tangled mess of ten nonlinear partial differential equations. They are incredibly hard to solve and for almost a century there have been many attempts at finding solutions which might describe real world phenomena. For the remainder of these lectures we are going to focus on one set of solutions which apply in a very particular regime. We will solve the Field Equations for the whole Universe under the assumption that it is homogeneous and isotropic.

For many centuries we have grown to believe that we don't live in a special place, that we are not at the center of the Universe. And, oddly enough, this point of view allows us to make some far reaching assumptions. So for example, if we are insignificant and, furthermore, everywhere is insignificant, then we can assume that at any given time, the Universe looks the same everywhere. In fact we can take that statement to an extreme and assume that at any given time, the Universe looks *exactly* the same at every single point in space. Such a space-time is dubbed to be *homogeneous*.

There is another assumption that takes into account the extreme regularity of the Universe and that is the fact that, at any given point in space, the Universe looks very much the same in whatever direction we look. Again such an assumption can be taken to an extreme so that at any point, the Universe look *exactly* the same, whatever direction one looks. Such a space time is dubbed to be *isotropic*.

Homogeneity and isotropy are distinct yet inter-related concepts. For example a universe which is isotropic will be homogeneous while a universe that is homogeneous *may not* be isotropic. A universe which is only isotropic around one point is not homogeneous. A universe that is both homogeneous and isotropic is said to satisfy the *Cosmological Principle*. It is believed that our Universe satisfies the Cosmological Principle.

Homogeneity severely restrict the metrics that we are allowed to work with in the Einstein field equation. For a start, they must be independent of space, and solely functions of time. Furthermore, we must restrict ourselves to spaces of constant curvature of which there are only three: a flat euclidean space, a positively curved space and a negatively curved space. We will look at curved spaces in a later lecture and will restrict ourselves to a flat geometry here.

The metric for a flat Universe takes the following form:

$$ds^2 = -c^2dt^2 + a^2(t)[(dx^1)^2 + (dx^2)^2 + (dx^3)^2]$$

We call $a(t)$ the scale factor and t is normally called *cosmic* time or *physical* time. The energy momentum tensor must also satisfy homogeneity and isotropy. If we consider a perfect fluid, we restrict ourselves to

$$T^{\alpha\beta} = (\rho + \frac{P}{c^2})U^\alpha U^\beta + Pg^{\alpha\beta}$$

with $U^\alpha = (c, 0, 0, 0)$ and ρ and P are simply functions of time. Note that both the metric and the energy-momentum tensor are diagonal. So

$$\begin{array}{ll} g_{00} = -1 & \qquad g_{ij} = a^2(t)\delta_{ij} \\ T_{00} = \rho c^2 & \qquad T_{ij} = a^2 P\delta_{ij} \end{array}$$

As we shall see, with this metric and energy-momentum tensor, the Einstein field equations are greatly simplified. We must first calculate the connection coefficients. We have that the only non-vanishing elements are (and from now one we will use $\dot{} = \frac{d}{dt}$ i.e. *not* to be confused with $\frac{d}{d\lambda}$ that we have used previously for the geodesic equations):

$$\begin{aligned} \Gamma^0{}_{ij} &= \frac{1}{c}a\dot{a}\delta_{ij} \\ \Gamma^i{}_{0j} &= \frac{1}{c}\frac{\dot{a}}{a}\delta^i_j \end{aligned}$$

and the resulting Ricci tensor is

$$\begin{aligned} R_{00} &= -\frac{3}{c^2}\frac{\ddot{a}}{a} \\ R_{0i} &= 0 \\ R_{ij} &= \frac{1}{c^2}(a\ddot{a} + 2\dot{a}^2)\delta_{ij} \end{aligned}$$

Again, the Ricci tensor is diagonal. We can calculate the Ricci scalar:

$$R = -R_{00} + \frac{1}{a^2}R_{ii} = \frac{1}{c^2}\left[6\frac{\ddot{a}}{a} + 6\left(\frac{\dot{a}}{a}\right)^2\right]$$

to find the two Einstein Field equations:

$$\begin{aligned} G_{00} = R_{00} - \frac{1}{2}Rg_{00} = \frac{8\pi G}{c^4}T_{00} \;\;&\Leftrightarrow\;\; 3\left(\frac{\dot{a}}{a}\right)^2 = 8\pi G\rho \\ G_{ij} = R_{ij} - \frac{1}{2}Rg_{ij} = \frac{8\pi G}{c^4}T_{ij} \;\;&\Leftrightarrow\;\; -2a\ddot{a} - \dot{a}^2 = \frac{8\pi G}{c^2}a^2P \end{aligned}$$

We can use the first equation to simplify the $2^{\rm nd}$ equation to

$$3\frac{\ddot{a}}{a} = -4\pi G(\rho + 3\frac{P}{c^2})$$

These two equations can be solved to find how the scale factor, $a(t)$, evolves as a function of time. The first equation is often known as the *Friedman-Robertson-Walker* equation or FRW equation and the metric is one of the three FRW metrics. The latter equation in $\ddot{a}$ is known as the *Raychauduri* equation.

Both of the evolution equations we have found are sourced by ρ and P. These quantities satisfy a conservation equation that arises from

$$\nabla_\alpha T^{\alpha\beta} = 0$$

and in the homogeneous and isotropic case becomes

$$\dot{\rho} + 3\frac{\dot{a}}{a}(\rho + \frac{P}{c^2}) = 0$$

It turns out that the FRW equation, the Raychauduri equation and the energy-momentum conservation equation are not independent. It is a straightforward exercise to show that you can obtain one from the other two. We are therefore left with two equations for three unknowns.

One has to decide what kind of energy we are considering and in a later lecture we will consider a variety of possibilities. But for now, we can hint at a substantial simplification. If we assume that the system satisfies an equation of state, so $P = P(\rho)$ and, furthermore that it is a *polytropic fluid* we have that

$$P = w\rho c^2 \tag{19}$$

where w is a constant, the *equation of state* of the system.

16 Properties of a Friedman Universe I

We can now explore the properties of these evolving Universes. Let us first do something very simple. Let us pick two objects (galaxies for example) that lie at a given distance from each other. At time t_1 they are at a distance r_1 while at a time t_2, they are at a distance r_2. We have that during that time interval, the change between r_1 and r_2 is given by

$$\frac{r_2}{r_1} = \frac{a(t_2)}{a(t_1)}$$

and, because of the cosmological principle, this is true *whatever* two points we would have chosen. It then makes to sense to parametrize the distance between the two points as

$$r(t) = a(t)x$$

where x is completely independent of t. We can see that we have already stumbled upon x when we wrote down the metric for a homogeneous and isotropic space time. It is the set of coordinates (x^1, x^2, x^3) that remain unchanged during the evolution of the Universe. We known that the real, *physical* coordinates are multiplied by $a(t)$ but (x^1, x^2, x^3) are time independent and are known as *conformal* coordinates. We can work out how quickly the two objects we considered are moving away from each other. We have that their relative velocity is given by

$$v = \dot{r} = \dot{a}x = \frac{\dot{a}}{a}ax = \frac{\dot{a}}{a}r \equiv Hr$$

In other words, the recession speed between two objects is proportional to the distance between them. This equality applied today (at t_0) is

$$v = H_0 r$$

and is known as *Hubble's Law* where H_0 is the Hubble constant and is given by $H_0 = 100h$ km s^{-1} Mpc^{-1} and h is a dimensionless constant which is approximately $h \simeq 0.7$.

How can we measure velocities in an expanding universe? Consider a photon with wavelength λ being emitted at one point and observed at some other point. We have that the Doppler shift is given by

$$\lambda' \simeq \lambda(1 + \frac{v}{c})$$

We can rewrite it in a differential form

$$\frac{d\lambda}{\lambda} \simeq \frac{dv}{c} = \frac{\dot{a}}{a}\frac{dr}{c} = \frac{\dot{a}}{a}dt = \frac{da}{a}$$

and integrate to find $\lambda \propto a$. We therefore have that wave lengths are stretched with the expansion of the Universe. It is convenient to define the factor by which the wavelength is stretched by

$$z = \frac{\lambda_r - \lambda_e}{\lambda_e} \to 1 + z \equiv \frac{a_0}{a}$$

where a_0 is the scale factor today (throughout these lecture notes we will choose a convention in which $a_0 = 1$). We call z the *redshift*.

For example, if you look at Figure 4 you can see the spectra measured from a galaxy; a few lines are clearly visible and identifiable. Measured in the laboratory on Earth (top panel), these lines will have a specific set of wavelengths but measured in a specific, distant, galaxy (bottom panel) the lines will be shifted to longer wavelengths. Hence a measurement of the redshift (or blueshift), i.e. a measurement of the Doppler shift, will be a direct measurement of the velocity of the galaxy.

The American astronomer, Edwin Hubble measured the distances to a number of distant galaxies and measured their recession velocities. The data he had was patchy, as you can see from Figure 5, but he was able to discern a a pattern: most of the galaxies are moving away from us *and* the further away they are, the faster they are moving. With more modern data, this phenomenon is striking, as you can see in the Figure 5. The data is neatly fit by a law of the form

$$v = H_0 r$$

where H_0 is a constant (known as Hubble's constant). Current measurements of this constant give us $H_0 = 67$ km s^{-1} Mpc^{-1}.

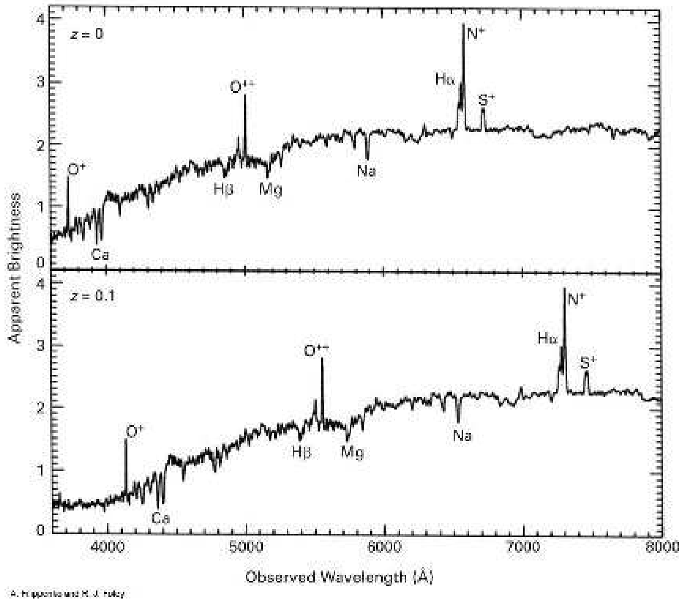

Figure 4: A set of spectra measured in laboratory (top panel) on a distant galaxy (bottom panel)

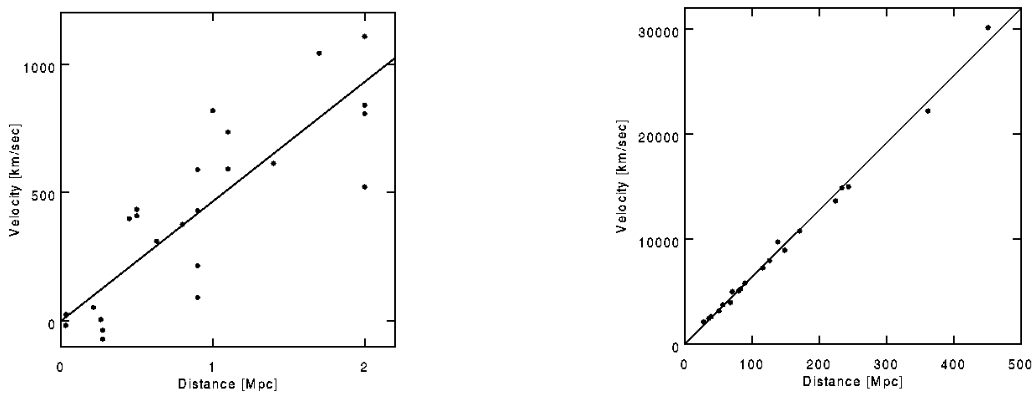

Figure 5: The recession velocity of galaxies, Hubble's data circa 1929 (left) and SN data circa 1995 (right)

17 Energy, Pressure and the History of the Universe

We can now solve the FRW equations for a range of different behaviours. In the final few lectures we will look, in some detail, at the nature of matter and energy in an expanding Universe but for now, we will restrict ourselves to describing them in terms of their equation of state in the form given in equation 19, $P = w\rho c^2$.

Let us start off with the case of *non-relativistic* matter. A notable example is that of massive particles whose energy is dominated by the rest energy of each individual particle. This kind of matter is sometimes simply called *matter* or *dust*. We can guess what the evolution of the mass density should be. The energy in a volume V is given by $E = Mc^2$ so $\rho c^2 = E/V$ where ρ is the mass density. But in an evolving Universe we have $V \propto a^3$ so $\rho \propto 1/a^3$. Alternatively, note that $P \simeq nk_BT \ll nMc^2 \simeq \rho c^2$ so $P \simeq 0$. Hence, using the conservation of energy equations:

$$\dot{\rho} + 3\frac{\dot{a}}{a}\rho = \frac{1}{a^3}\frac{d}{dt}(\rho a^3) = 0$$

and solving this equation we find $\rho \propto a^{-3}$. We can now solve the FRW equation (taking $\rho(a=1)\rho_0$):

$$\begin{aligned}\left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\frac{\rho_0}{a^3} \\ a^{1/2}\dot{a} &= \left(\frac{8\pi G\rho_0}{3}\right)^{1/2}\end{aligned}$$

to find $a \propto t^{2/3}$. If $a(t_0) = 1$, where t_0 is the time today, we have $a = (t/t_0)^{2/3}$ You will notice a few things. First of all, at $t = 0$ we have $a = 0$ i.e. there is an initial *singularity* known as the *Big Bang*. Furthermore we have that $v = \frac{\dot{a}}{a}r = \frac{2}{3t_0}r$. i.e. by measuring Hubble's law we measure the age of the Universe. And finally we have that $\ddot{a} < 0$ i.e. the Universe is *decelerating*.

The case of *Relativistic Matter* encompasses particles which are massless like photons or neutrinos. Recall that their energy is given by $E = h\nu = h2\pi/\lambda$ where ν is the frequency and λ is the wavelength. As we saw in Section 16, wavelengths are redshifted, i.e. $\lambda \propto a$ and hence the energy of an individual particle will evolve as $E \propto 1/a$. Once again, the mass density is given by $\rho c^2 = E/V \propto 1/(V\lambda) \propto 1/(a^3 a) = 1/a^4$. So the energy density of radiation decreases far more quickly than that of dust. We can go another route, using the equation of state and conservation of energy. We have that for radiation $P = \rho c^2/3$, so

$$\dot{\rho} + 4\frac{\dot{a}}{a}\rho = \frac{1}{a^4}\frac{d}{dt}(\rho a^4) = 0$$

which can be solved to give $\rho \propto a^{-4}$. We can now solve the FRW equations.

$$\begin{aligned}\left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\frac{\rho_0}{a^4} \\ a\dot{a} &= \left(\frac{8\pi G\rho_0}{3}\right)^{1/2}\end{aligned}$$

to find $a \propto t^{1/2}$ or $a = (t/t_0)^{1/2}$ Once again, the universe is decelerating but now $H_0 = 1/(2t_0)$. Note that there is a different relation between H_0 and t_0 so if we are to infer the age of the Universe from the expansion rate, we need to know what it contains.

It is straightforward to consider a general w. Energy-momentum conservation gives us

$$\dot{\rho} + 3(1+w)\frac{\dot{a}}{a}\rho = \frac{1}{a^{3(1+w)}}\frac{d}{dt}(\rho a^{3(1+w)}) = 0$$

Note that as w gets smaller and more negative, ρ decays more slowly. We can solve the FRW equations

$$\begin{aligned}\left(\frac{\dot{a}}{a}\right)^2 &= \frac{8\pi G}{3}\frac{\rho_0}{a^{3(1+w)}} \\ a^{(1+3w)/2}\dot{a} &= \left(\frac{8\pi G\rho_0}{3}\right)^{1/2}\end{aligned}$$

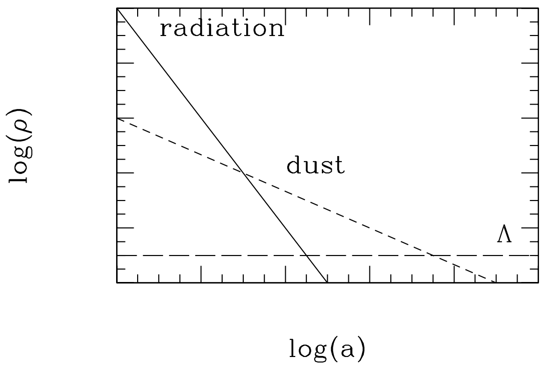

Figure 6: The energy density of radiation, matter and the cosmological constant as a function of time

to find $a = (t/t_0)^{2/3(1+w)}$ which is valid if $w > -1$. For $w < -1/3$ the expansion rate is *accelerating*, not decelerating. For the special case of $w = -1/3$ we have $a \propto t$.

Finally, we should consider the very special case of a *Cosmological Constant*. Such odd situation arises in the extreme case of $P = -\rho c^2$. You may find that such an equation of state is obeyed by vaccum fluctuations of matter. Such type of matter can be described by the Λ we found in equation 18. The solutions are straightforward: ρ is constant, $\frac{\dot{a}}{a}$ is constant and $a \propto \exp(Ht)$.

Throughout this section, we have considered one type of matter at a time but it would make more sense to consider a mix. For example we know that there are photons *and* protons in the Universe so in the very least we need to include both types of energy density in the FRW equations:

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\left(\frac{\rho_{M0}}{a^3} + \frac{\rho_{R0}}{a^4}\right)$$

In fact, the current picture of the universe involves all three types of matter/energy we considered in this section and, depending on their evolution as a function of a, they will dominate the dynamics of the Universe at different times. In Figure 6 we plot the energy densities as a function of scale factor and we can clearly see the three stages in the Universe's evolution: a *radiation era*, followed by a *matter era* ending up with a *cosmological constant era* more commonly known as a Λ *era*.

18 Geometry and Destiny

Until now we have restricted ourselves to a flat Universe with Euclidean geometry. Before we move away from such spaces let us revisit the metric. We have

$$ds^2 = -c^2dt^2 + a^2(t)(dx^2 + dy^2 + dz^2)$$

Let us transform to spherical polar coordinates

$$x = r\cos\phi\sin\theta$$

$$\begin{aligned} y &= r\sin\phi\sin\theta \\ z &= r\cos\theta \end{aligned}$$

and rewrite the metric

$$ds^2 = -c^2dt^2 + a^2(t)(dr^2 + r^2d\theta^2 + r^2\sin^2\theta d\phi^2)$$

We could in principle work out the FRW and Raychauduri equations in this coordinate system.

Let us now consider a 3 dimensional surface that is positively curved. In other words, it is the surface of a 3 dimensional hypersphere in a fictitious space with 4 dimensions. The equation for the surface of a sphere in this 4 dimensional space, with coordinates (X, Y, Z, W), is

$$X^2 + Y^2 + Z^2 + W^2 = R^2$$

Now in the same way that we can construct spherical coordinates in three dimensions, we can build hyperspherical coordinates in 4 dimensions:

$$\begin{aligned} X &= R\sin\chi\sin\theta\cos\phi \\ Y &= R\sin\chi\sin\theta\sin\phi \\ Z &= R\sin\chi\cos\theta \\ W &= R\cos\chi \end{aligned}$$

We can now work out the line element on the surface of this hyper-sphere

$$ds^2 = dX^2 + dY^2 + dZ^2 + dW^2 = R^2\left[d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\right]$$

Note how different it is from the flat geometry. If we transform $R\sin\chi$ into r for it all to agree we have that

$$d\chi^2 = \frac{dr^2}{R^2 - r^2}$$

We can now repeat this exercise for 3-D surface with negative curvature- a hyper-hyperboloide so to speak. In our fictitious 4-D space (not to be confused with space time), we have that the surface is defined by

$$X^2 + Y^2 + Z^2 - W^2 = -R^2$$

Let us now change to a good coordinate system for that surface:

$$\begin{aligned} X &= R\sinh\chi\sin\theta\cos\phi \\ Y &= R\sinh\chi\sin\theta\sin\phi \\ Z &= R\sinh\chi\cos\theta \\ W &= R\cosh\chi \end{aligned}$$

The line element on that surface will now be

$$ds^2 = dX^2 + dY^2 + dZ^2 + dW^2 = R^2\left[d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta d\phi^2)\right]$$

We can replace $R\sinh\chi$ by r to get

$$d\chi^2 = \frac{dr^2}{R^2+r^2}$$

We can clearly write all three space time metrics (flat, hyperspherical, hyper-hyperbolic) in a unified way. If we take $r = R\sin\chi$ for the positively curved space and $r = R\sinh\chi$ for the negatively curved space we have

$$ds^2 = -c^2dt^2 + a^2(t)\left[\frac{d^2r}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2)\right] \tag{20}$$

where k is positive, zero or negative for spherical, flat or hyperbolic geometries, and $|k| = 1/R^2$.

We can now repeat the calculation we undertook for a flat geometry and find the connection coefficients, Ricci tensor and scalar and the evolution equations. Take the metric $g_{\alpha\beta} = diag(-1, \frac{a^2}{1-kr^2}, a^2r^2, a^2r^2\sin\theta^2)$ and note that for this choice of coordinates, the i and j labels now run over r, θ and ϕ. We find that the connection coefficients are:

$$\begin{aligned}
\Gamma^0_{ij} &= \frac{1}{c}a\dot{a}\tilde{g}_{ij} \\
\Gamma^i_{0j} &= \frac{1}{c}\frac{\dot{a}}{a}\delta^i_j \\
\Gamma^i_{jk} &= \tilde{\Gamma}^i_{jk}
\end{aligned}$$

where $\tilde{g}_{ij}$ and $\tilde{\Gamma}$ are the metric and connection coefficients of the conformal 3-space (that is of the 3-space with the conformal factor, a, divided out):

$$\begin{aligned}
\tilde{\Gamma}^r{}_{rr} &= \frac{kr}{1-kr^2} \\
\tilde{\Gamma}^r{}_{\theta\theta} &= -r(1-kr^2) \\
\tilde{\Gamma}^r{}_{\phi\phi} &= -(1-kr^2)r\sin^2(\theta) \\
\tilde{\Gamma}^\theta{}_{\theta r} &= \frac{1}{r} \\
\tilde{\Gamma}^\theta{}_{\phi\phi} &= \frac{-\sin(2\theta)}{2} \\
\tilde{\Gamma}^\phi{}_{\phi r} &= \frac{1}{r} \\
\tilde{\Gamma}^\phi{}_{\theta\phi} &= \frac{1}{\tan(\theta)}
\end{aligned}$$

The Ricci tensor and scalar can be combined to form the Einstein tensor

$$\begin{aligned}
G_{00} &= 3\frac{\dot{a}^2+kc^2}{c^2a^2} \\
G_{ij} &= -\frac{2a\ddot{a}+\dot{a}^2+kc^2}{c^2}\tilde{g}_{ij}
\end{aligned} \tag{21}$$

while the energy-momentum tensor is

$$\begin{aligned} T_{00} &= \rho c^2 \\ T_{ij} &= a^2 P \tilde{g}_{ij} \end{aligned}$$

Combining them gives us the Friedman equation

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho - \frac{kc^2}{a^2}$$

while the Raychauduri equations remains as

$$3\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}(\rho + 3\frac{P}{c^2})$$

Let us now explore the consequences of the overall geometry of the Universe, i.e. the term proportional to k in the FRW equations: For simplicity, let us consider a dust filled universe. We can see that the term proportional to k will only be important at late times, when it dominates over the energy density of dust. In other words, in the universe we can say that *curvature dominates* at late times. Let us now consider the two possibilities. First of all, let us take $k < 0$. We then have that

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho + \frac{|k|c^2}{a^2}$$

When the curvature dominates we have that

$$\left(\frac{\dot{a}}{a}\right)^2 = \frac{|k|c^2}{a^2}$$

so $a \propto t$. In this case, the scale factor grows at the speed of light. We can also consider $k > 0$. From the FRW equations we see that there is a point, when $\frac{8\pi G}{3}\rho = \frac{kc^2}{a^2}$ and therefore $\dot{a} = 0$ when the Universe stops expanding. At this point the Universe starts contracting and evolves to a *Big Crunch*. Clearly geometry is intimately tied to destiny. If we know the geometry of the Universe we know its future.

There is another way we can fathom the future of the Universe. If $k = 0$, there is a strict relationship between $H = \frac{\dot{a}}{a}$ and ρ. Indeed from the FRW equation we have

$$H^2 = \left(\frac{\dot{a}}{a}\right)^2 = \frac{8\pi G}{3}\rho \to \rho = \rho_c \equiv \frac{3H^2}{8\pi G}$$

We call ρ_c the *critical density*. It is a function of a. If we take $H_0 = 100h\text{Km s}^{-1}\text{ Mpc}^{-1}$, we have that

$$\rho_c = 1.9 \times 10^{-26} h^2 \text{kgm}^{-3}$$

which corresponds to a few atoms of Hydrogen per cubic meter. Compare this with the density of water which is 10^3 kg m^{-3}. Now let us take another look at the FRW equation and rewrite it as

$$\frac{1}{2}\dot{a}^2 - \frac{4\pi G}{3}\rho a^2 = -\frac{1}{2}kc^2$$

which has the form $E_{tot} = U + K$ and we equate E_{tot} to $-kc^2$ so that K is the kinetic energy, U is the gravitational energy. We see that if $\rho = \rho_c$, it corresponds to the total energy of the system being 0, i.e. kinetic and gravitational energy balance themselves out perfectly. Let us look at the case of nonzero k.

$k < 0$ $\rho < \rho_c$ and therefore total energy is positive, kinetic energy wins out and the Universe expands at a constant speed.

$k > 0$ $\rho > \rho_c$ and the total energy is negative, gravitational energy wins out and the Universe recollapses.

We recover an important underlying principle behind all this, the geometry is related to the energy density.

It is convenient to define a more compact notation. The *fractional energy density* or *density parameter*. We define

$$\Omega \equiv \frac{\rho}{\rho_c}$$

It will be a function of a and we normally express its value today as Ω_0. If there are various contributions to the energy density, we can define the fractional energy densities of each one of these contributions. For example

$$\Omega_R \equiv \frac{\rho_R}{\rho_c} \qquad \Omega_M \equiv \frac{\rho_M}{\rho_c} \qquad \cdots$$

It is convenient to define two additional Ωs:

$$\begin{aligned} \Omega_\Lambda &\equiv \frac{\Lambda}{3H^2} \\ \Omega_k &\equiv -\frac{kc^2}{a^2H^2} \end{aligned}$$

and we have Ω:

$$\Omega = \Omega_R + \Omega_M + \Omega_\Lambda$$

We now have

$\Omega < 1$: $\rho < \rho_c$, $k < 0$, Universe is open (hyperbolic)

$\Omega = 1$: $\rho = \rho_c$, $k = 0$, Universe is flat (Euclidean)

$\Omega > 1$: $\rho > \rho_c$, $k > 0$, Universe is closed (spherical)

If we divide the FRW equation through by ρ_c we find that it can be rewritten as

$$H^2(a) = H_0^2 \left[\frac{\Omega_{M0}}{a^3} + \frac{\Omega_{R0}}{a^4} + \frac{\Omega_{K0}}{a^2} + \Omega_\Lambda \right] \tag{22}$$

where the subscript "0" indicates that these quantities are evaluated at t_0. We will normally drop the subscript when referring to the various Ωs evaluated today. When we refer to the Ωs at different times, we will explicitely say so or add an argument (for example $\Omega_M(a)$ or $\Omega_M(z)$).

How does Ω evolve? Without loss of generality, let us consider a Universe with dust, take the FRW equations and divide by H^2 to obtain

$$\Omega - 1 = \frac{kc^2}{a^2 H^2} \propto k t^{2/3}$$

I.e., if $\Omega \neq 1$, it is unstable and driven away from 1. The same is true in a radiation dominated universe and for any *decelerating* Universe: $\Omega = 1$ is an unstable fixed point and, as we saw above, curvature dominates at late times.

19 Properties of a Friedman Universe I

Let us revisit the properties of a FRW universe, now that we know a bit more about the the evolution of the scale factor. Distances play an important role if we are to map out its behaviour in detail. We have already been exposed to Hubble's law

$$v = H_0 d$$

from which we can extract Hubble's constant. From Hubble's constant we can define a *Hubble time*

$$t_H = \frac{1}{H_0} = 9.78 \times 10^9 \ h^{-1} \ \mathrm{yr}$$

and the *Hubble distance*

$$D_H = \frac{c}{H_0} = 3000 \ h^{-1} \ \mathrm{Mpc}$$

These quantities set the scale of the Universe and give us a rough idea of how old it is and how far we can see. They are only rough estimates and to get a firmer idea of distances and ages, we need to work with the metric and FRW equations more carefully.

To actually figure out how far we can see, we need to work out how far a light ray travels over a given period of time. To be specific, what is the distance, D_M to a galaxy that emitted a light ray at time t, which reaches us today? Let us look at the expression for the metric used in equation 20 for a light ray. We have that

$$\frac{dr^2}{1 - kr^2} = \frac{c^2 dt^2}{a^2(t)} \tag{23}$$

The time integral gives us the *comoving distance*:

$$D_C = c \int_t^{t_0} \frac{dt'}{a(t')}$$

From equation 22 we have that $-k = \Omega_k / D_H^2$. Performing the radial integral (and assuming the observer is at $r = 0$ we have

$$\int_0^{D_M} \frac{dr}{\sqrt{1 - kr^2}} = \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sinh^{-1}[\sqrt{\Omega_k} D_M / D_H] & \text{for } \ \Omega_k > 0 \\ D_M & \text{for } \ \Omega_k = 0 \\ \frac{D_H}{\sqrt{|\Omega_k|}} \sin^{-1}[\sqrt{|\Omega_k|} D_M / D_H] & \text{for } \ \Omega_k < 0 \end{cases}$$

so we find an expression for the *proper motion distance* (also known as the *transverse comoving distance*, D_M in terms of the comoving distance)

$$D_M = \begin{cases} \frac{D_H}{\sqrt{\Omega_k}} \sinh[\sqrt{\Omega_k} D_C/D_H] & \text{for } \Omega_k > 0 \\ D_C & \text{for } \Omega_k = 0 \\ \frac{D_H}{\sqrt{|\Omega_k|}} \sin[\sqrt{|\Omega_k|} D_C/D_H] & \text{for } \Omega_k < 0 \end{cases}$$

Suppose now we we look at an object of a finite size which is transverse to our line of sight and lies at a certain distance from us. If we divide the physical transverse size of the object by the angle that object subtends in the sky (the angular size of the object) we obtain the *angular diameter distance*:

$$D_A = \frac{D_M}{1+z}$$

Hence, if we know that size of an object and its redshift we can work out, for a given Universe, D_A.

Alternatively, we may know the brightness or luminosity of an object at a given distance. We know that the flux of that object at a distance D_L is given by

$$F = \frac{L}{4\pi D_L^2}$$

D_L is aptly known as the *luminosity distance* and is related to other distances through:

$$D_L = (1+z)D_M = (1+z)^2 D_A$$

It turns out that, in astronomy, one often works with a logarithmic scale, i.e. with *magnitudes*. One can define the *distance modulus*:

$$DM \equiv 5 \log\left(\frac{D_L}{10 \text{ pc}}\right)$$

and it can be measured from the *apparent magnitude* m (related to the flux at the observer) and the *absolute magnitude* M (what it would be if the observer was at 10 pc from the source) through

$$m = M + DM$$

We now have a plethora of distances which can be deployed in a range of different observations. They clearly depend on the universe we are considering, i.e. on the values of H_0, and the various Ωs. While Ω_k will dictate the geometry, D_c will depend on how the Universe evolves. It is useful to rewrite D_c in a few different ways. It is useful to use the FRW in the form presented in equation 22. We can transform the time integral in D_c to an integral in a:

$$D_C = \int_t^{t_0} \frac{cdt'}{a(t')} = c\int_a^1 \frac{da}{a^2 H(a)} = D_H \int_a^1 \frac{da}{a^2\sqrt{\Omega_M/a^3 + \Omega_R/a^4 + \Omega_k/a^2 + \Omega_\Lambda}}$$

An interesting question is how far has light travelled, from the big bang until now? This is known as the *particle horizon*, r_P and a naive estimate would be $r_P \simeq ct_0$ but that doesn't take into account the expansion of space time. The correct expression is given above and it is

$$r_P = D_C(0)$$

where the argument implies that it is evaluated from $t = 0$ to $t = t_0$. Applying it now to the simple case of a dust filled, flat Universe. We have that

$$r_P = 3ct_0$$

Unsurprisingly, the expansion leads to an extra factor.

We could ask a different question: how far can light travel from now until the infinite future, i.e. how much will we ever see of the current Universe. Known as the *event horizon* it is by the integral of equation 23 from t_0 until ∞. For example in a flat Universe we have

$$r_E = \int_{t_0}^{\infty} \frac{cdt'}{a(t')}$$

For a dust or radiation dominated universe we have that $r_E = \infty$ but this is not so for a universe dominated by a cosmological constant.

We have been focusing on distances but we can also improve our estimate of ages. We defined the Hubble time above and that is a rough estimate of the age of the Universe. To do better we need to resort to the FRW equations again, as above we have that $\dot{a} = aH$ so

$$dt = \frac{da}{aH} \rightarrow \int_0^{t_0} dt = \int_0^1 \frac{da}{aH} = t_0$$

which, combined with equation 22 gives us

$$t_0 = \frac{1}{H_0} \int_0^1 \frac{da}{a\sqrt{\Omega_M/a^3 + \Omega_R/a^4 + \Omega_k/a^2 + \Omega_\Lambda}}$$

We can use the above equation quite easily. For a flat, dust dominated Universe we find $t_0 = 2/(3H_0)$. If we now include a cosmological constant as well, we find

$$t_0 = H_0^{-1} \int_0^1 \frac{da}{a\sqrt{\Omega_M/a^3 + \Omega_\Lambda}}$$

At $\Omega_\Lambda = 0$ we simply retrieve the matter dominated result, but the larger Ω_Λ is, the older the Universe. To understand why, recall the Raychauduri equation for this Universe:

$$\frac{\ddot{a}}{a} = -\frac{4\pi G}{3}\rho + \frac{\Lambda}{3}$$

Divide by H_0^2 and we have that the *deceleration parameter*

$$q_0 \equiv -\frac{a(t_0)\ddot{a}(t_0)}{\dot{a}^2(t_0)} = \frac{1}{2}\Omega_M - \Omega_\Lambda$$

If $\Omega_M + \Omega_\Lambda = 1$ then $q_0 = \frac{3}{2}\Omega_M - 1$. If $\Omega_M < \frac{2}{3}$ we have $q_0 < 0$ and the Universe is accelerating. This lets us understand why the Universe is older. Take a $\Omega_\Lambda = 0$ and a $\Omega_\Lambda > 0$ which both have the *same* expansion rate today. The latter is accelerating which means it was expanding more slowly in the past than the former. This means it must have taken longer to reach its current speed and hence is older. Furthermore we can see that our inference about the Universe depends on our knowledge of the various Ωs. In other words, if we want to measure the age of the Universe we must also measure the density in its various components.

Finally, let us revisit Hubble's law. We worked out the relationship between velocities and distance for two objects which were very close to each other. If we want to consider objects which are further apart (not too distant galaxies) we can Taylor expand the scale factor today, we find that

$$a(t) = a(t_0) + \dot{a}(t_0)[t - t_0] + \frac{1}{2}\ddot{a}(t_0)[t - t_0]^2 + \cdots$$

Assume that the distance to the emitter at time t is roughly given by $d =\simeq c(t_0 - t)$ we can rewrite it as

$$(1 + z)^{-1} = 1 - H_0\frac{d}{c} - \frac{q_0 H_0^2}{2}\left(\frac{d}{c}\right)^2 + \cdots$$

For $q_0 = 0$ and small z we recover the Hubble law, $cz = H_0 d$. As we go to higher redshift, this is manifestly not good enough.

20 The Cosmological Distance ladder

Given our model of a range of possible universes, we would like to pin down which set of *cosmological parameters* (like t_0, H_0, Ω_M, $\cdots$) correspond to our Universe. We can ask questions like: what is the age of the Universe, is it accelerating or decelerating, what is its density and geometry? Interestingly enough, all these questions must be answered together and to do so we need to go out, observe and measure.

The first step is to map out the Universe and measure distances and redshifts accurately. By far the easiest quantity to measure is the redshift. By looking at the shift in the spectra of known elements it is possible to infer the recession velocity of the galaxy directly. Measuring distances is much harder. The most direct method is to use parallax to measure the distance to a star. Let us remember what you do here. Imagine that you look at an object in the sky. It can be described in terms of two angles. It has a position on the celestial sphere. Now imagine that we move a distance $2d$ from where we were. The object may move an angle θ from where it was. The angle that it has moved will be related to the distance D and displacement d. If we say $\theta = 2\alpha$ then we have $\tan\alpha = \frac{d}{D}$ If α is small then we can use the small angle approximation to get

$$\alpha = \frac{d}{D}$$

The motion of the earth around the sun gives us a very good baseline with which to measure distance. The distance from the earth to the sun is 1 AU so we have that $D = \frac{1}{\alpha}$ where α is

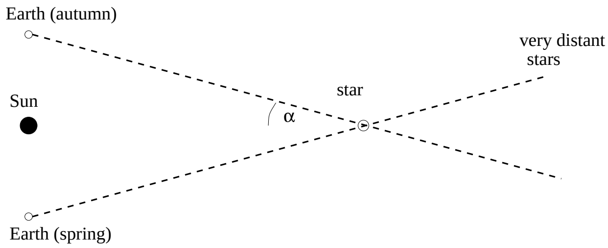

Figure 7: The motion of the Earth around the Sun supplies us with a long baseline for parallax measurments.

measured in arcseconds. D is then given in *parsecs.* One parsec corresponds to 206,265 AU or 3.09×10^{13}km. This is a tremendous distance, $1pc \sim 3.26$ light years. All stars have parallax angles less than one arcsecond. The closest star, Proxima Centauri, has a distance of 1.3pc. In 1989 a satellite was launched called Hipparcos to measure the distances to 118,000 stars with an accuracy of 0.001 arcseconds. This corresponds to distances of hundreds of parsecs. This may seem far but it isn't. The sun is 8kpc away from the centre of the galaxy.

We would like to be able to look further. The basic tool for doing this is to take an object of known brightness and see how bright it looks. Take a star with a given luminosity L. The luminosity is the amount of light it pumps out per second. How bright will it look from where we stand? We can think of standing on a point of a sphere of radius D centred on the star. The brightness will be $B = \frac{L}{4\pi D^2}$ The further away it is the dimmer it will look. If we know the luminosity of a star and we measure its brightness, then we will know how far away it is.

How can we do that in practice? Stars have varying luminosities and are very different. Is there any way in which we can use information about a star's structure to work out it's luminosity? Let us start by looking at the colours of stars. Different stars will emit different spectra. Some will look redder, others more yellow, while others will be blue. Their colours (or spectra) are intimately tied to their temperature. Remember a black body what black body looks like. Its spectrum peaks at a certain value which is given by its temperature. For example, the Sun is yellow-white, has a temperature of 5800 K. The star Bellatra is blue and has a temperature of 21,500 K. Betelgeuse is red and has a temperature of 3500 K. Now we might think that we have it made.

The luminosity must be related to the temperature somehow. If we assume that it is black body, the energy flux is $F = \sigma T^4$ where σ is the Stefan-Boltzmann constant $\sigma = 5.6\times10^{-8}Wm^{-2}K^{-4}$. So luminosity is simply the surface of the star times its flux $L = 4\pi R^2\sigma T^4$. There is indeed a very tight connection *but* stars can have different radii. For example main sequence stars have one type of radius while red giants have much larger radii. We can look at the H.R. diagram and find stars with the same temperature which have very different luminosities. However if we can identify what type of stars they are then we can, given their colours, read off their luminosities.

Suppose we look at the spectra of two stars, A and B, and we identify some spectral lines. These correspond to the same absorption/emission lines but in A they're narrower than in B. What leads to the thickness of the lines? If there are random velocities, they will Doppler shift the line. The larger the spread in velocities, the more shifts there will be. But clearly for there to be a larger spread, they have to be closer to the core of the star i.e. the radius

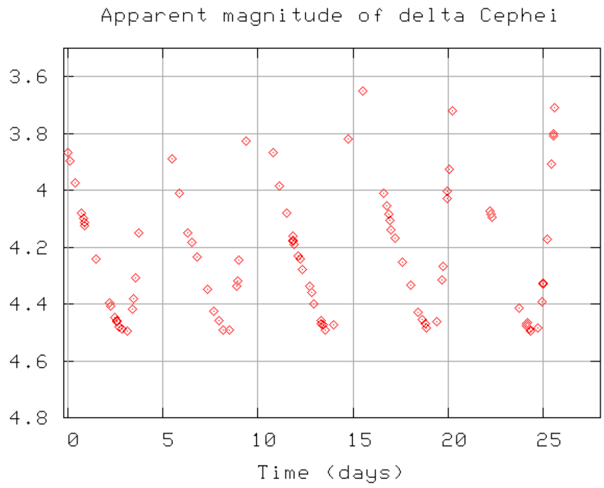

Figure 8: The luminosity of Cepheid stars varies periodically over time.

has to be smaller. I.e. broader lines imply smaller R. So by reading off the thickness of the lines we can pinpoint what type of stars they are and then from their colours we can infer their luminosity. For example: Sun has $T \sim 5800K$. It is a main sequence star with a luminosity of $1\,\mathrm{L}_\odot$. Aldebaran is a giant star which, even though it is cooler, $T \sim 4000K$, has a luminosity of $370\,\mathrm{L}_\odot$. This method, known as *spectroscopic parallax* can be used to go out to 10kpc.

How can we move out beyond 10kpc? There are some stars which have a very useful property. Their brightness varies with time and the longer their variation, the larger their luminosity. These stars known as *Cepheid* stars are interesting because they have a) periods of days (which means their variations can be easily observed) and b) are very luminous with luminosites of about $100 - 1000\,\mathrm{L}_\odot$ and therefore they can be seen at great distances. It was found that their period of oscillation is directly related to their intrinsic luminosity.

These stars pulsate because their surface oscillates up and down like a spring. The gas of the star heats up and then cools down, and the interplay of pressure and gravity keeps it pulsating. How do we know the intrinsic luminosity of these stars? We pick out globular clusters (very bright aglomerations in the Galaxy with about 10^6 stars) and we use spectroscopic parallax to measure their distances. Then we look for the varying stars, measure their brightness and period and build up a plot. There is another class of star called RR Lyrae which also oscillate. They have much shorter periods, and are less luminous ($\sim 100\,\mathrm{L}_\odot$) but have a much tighter relationship between period and luminosity. We can use these stars to go out to 30Mpc (i.e. 30 million parsecs). If we want to go further, we need something which is even brighter.

The method of choice for measuring very large distances is to look for distant *supernovae*. As you know, supernovae are the end point of stellar evolution, massive explosions that pump out an incredible amount of energy. Indeed supernovae can be as luminous as the galaxies which host them with luminosities of around $10^9\,\mathrm{L}_\odot$. So we can see distant supernovae, measure their brightness and if we know their luminosities, use the inverse square law to measure the distance. A certain type of supernova (supernovae I_a) seem to have very similar behaviours. They don't all have the same luminosities but the rate at which they fade after explosion is intimately tied

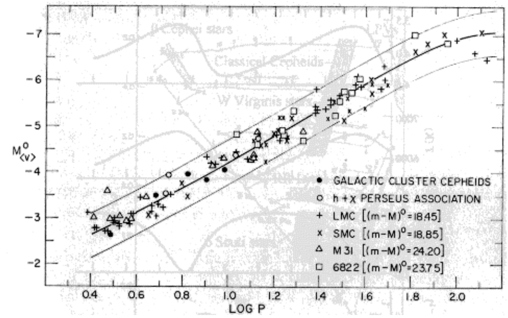

Figure 9: There is a tight relationship between the period (x-axis) and the luminosity (or magnitude in the y-axis) for Cepheid and RRLyrae star.

to the luminosity at the moment of the explosion. So by following the ramp up to the explosion and the subsequent decay it is possible to recalibrate a supernova explosion so that we know its luminosity.

Supernovae *Ia* arise when a white dwarf which is just marginally heavier than the Chandrasekhar mass gobbles up enough material to become unstable and collapse. The electron degeneracy pressure is unable to hold it up and it collapses in a fiery explosion. Supernovae can be used to measure distances out to a distance of about 1000Mpc. They are extremely rare, one per galaxy per hundred years, so we have to be lucky to find them. However there are 10^9 galaxies to look at so the current practice is to stare at large concentrations of galaxies and wait for an event to erupt. Of order 500 SN have been measured in the past decades.

Finally I want to mention another distance indicator which can be used to measure the distances out to about 100 Mpc. When we look at distant galaxies there is a very useful spectral line to measure. It has a wavelength of 21cm and corresponds to the energy associated with the coupling of the spin of the nucleus (a proton) with the spin of an electron in a Hydrogen atom. If they are aligned, the energy will be higher than if they are anti-aligned. Once again, this line will have a certain width due to the Doppler effect as a result of internal motions in the galaxy. In particular the rotation of the hydrogen will induce a Doppler effect. The faster the rotation, the larger the Doppler effect and the wider the spectral line.

We know, from Newtonian gravity that the rotation is intimately tied to the mass of the galaxy, so the wider the line, the faster the speed of rotation and hence the more massive the galaxy. But the more massive the galaxy, the more stars it should contain and therefore the more luminous it should be. So by measuring the 21cm line it is possible to measure the luminosity of distant galaxies. This is known as the *Tully Fisher* relation.

We can now use these techniques to pin down various properties of our Universe.

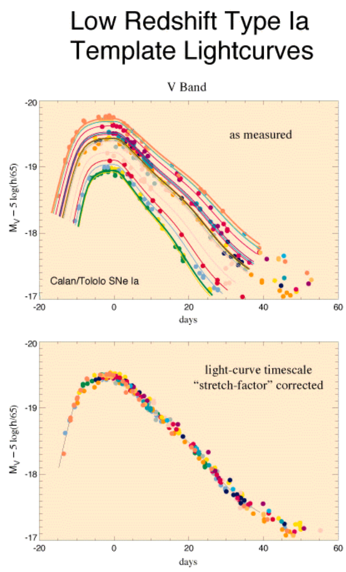

Figure 10: Top panel: the light curves of Supernovae Ia. Bottom panel: the light curves have been recalibrated (or “stretched”) so that they all have the same decay rate. Note that, following this procedure, all curves have the same luminosity at the peak.

21 The Thermal history of the Universe: Equilbrium

We shall now look at how the contents of the Universe are affected by expansion. The first property which we must consider is that as the Universe expands, its contents cool down. How can we see that? Let us focus on the radiation contained in the Universe. In the previous sections we found that the energy density in radiation decreases as

$$\rho \propto \frac{1}{a^4}.$$

What else can we say about radiation? Let us make a simplifying assumption, that it is in thermal equilibrium and therefore behaves like a blackbody. For this to be true, the radiation must interact very efficiently with itself to redistribute any fluctuations in energy and occupy the maximum entropy state. You have studied the properties of radiation (or relativistic particles) in thermal equilibrium in statistical mechanics in the 2^{nd} year and found that it can be described in terms of an *occupation number per mode* given by

$$F(\nu) = \frac{2}{\exp \frac{h\nu}{k_B T} - 1}$$

where ν is the frequency of the photon. This corresponds to an *energy density per mode*

$$\epsilon(\nu) d\nu = \frac{8\pi\nu^3 d\nu}{c^3} \frac{h}{\exp \frac{h\nu}{k_B T} - 1}$$

If we integrate over all frequencies we have that the energy density in photons is:

$$\rho_\gamma = \frac{\pi^2}{15}(k_B T)\left(\frac{k_B T}{\hbar c}\right)^3 . \qquad (24)$$

We have therefore that $\rho_\gamma \propto T^4$. Hence if radiation is in thermal equilibrium we have that

$$T \propto \frac{1}{a}$$

Is this the temperature of the Universe? Two ingredients are necessary. First of all, everything else has to feel that temperature which means they have to interact (even if only indirectly) with the photons. For example the scattering off photons of electrons and positrons is through the emission and absorption of photons. And once again, at sufficiently high temperatures, everything interacts quite strongly.

Another essential ingredient is that the radiation must dominate over the remaining forms of matter in the Universe. We have to be careful with this because we know that different types of energy will evolve in different ways as the Universe expands. For example we have that the energy density of dust (or non-relativistic matter) evolves as $\rho_{NR} \propto a^{-3}$ as compared to $\rho_\gamma \propto a^{-4}$ so even if ρ_γ was dominant at early times it may be negligible today. However we also know that the *number density* of photons, $n_\gamma \propto a^{-3}$ as does the number density of non-relativistic particles, $n_{NR} \propto a^{-3}$. If we add up all the non-relativisitic particle in the form of neutrons and protons (which we call baryons), we find that number density of baryons, n_B is very small compared to the number density of photons. In fact we can define the *baryon to entropy ratio*, η_B:

$$\eta_B = \frac{n_B}{n_\gamma} \simeq 10^{-10}$$

As we can see there are many more photons in the Universe than particles like protons and neutrons. So it is safe to say that the temperature of the photons sets the temperature of the Universe.

We can think of the Universe as a gigantic heat bath which is cooling with time. The temperature decreases as the inverse of the scale factor. To study the evolution of matter in the Universe we must now use statistical mechanics to follow the evolution of the various components as the temperature decreases. Let us start off with an ideal gas of bosons or fermions. Its occupation number per mode (now labeled in terms of momentum $\mathbf{p}$) is

$$F(\mathbf{p}) = \frac{g}{\exp\left(\frac{E-\mu}{k_B T}\right) \pm 1}$$

where g is the degeneracy factor, $E = \sqrt{p^2c^2 + M^2c^4}$ is the energy, μ is the chemical potential and + (-) corresponds to the Fermi-Dirac (Bose-Einstein) distribution. We can use this expression to calculate some macroscopic quantities such as the number density

$$n = \frac{g}{h^3}\int \frac{d^3p}{\exp\left(\frac{E-\mu}{k_B T}\right) \pm 1}$$

the energy density

$$\rho c^2 = \frac{g}{h^3}\int \frac{E(\mathbf{p})d^3p}{\exp\left(\frac{E-\mu}{k_BT}\right)\pm 1}$$

and the pressure

$$P = \frac{g}{h^3}\int \frac{p^2c^2}{3E}\frac{d^3p}{\exp\left(\frac{E-\mu}{T}\right)\pm 1}$$

It is instructive to consider two limits. First of all let us take the case where the temperature of the Universe corresponds to energies which are much larger than the rest mass of the individual particles, i.e. $k_BT \gg Mc^2$ and let us take $\mu \simeq 0$. We then have that the number density obeys

$$\begin{aligned} n &= \frac{\zeta(3)}{\pi^2}g\left(\frac{k_BT}{\hbar c}\right)^3 \quad \text{(B.E.)} \\ n &= \frac{3\zeta(3)}{4\pi^2}g\left(\frac{k_BT}{\hbar c}\right)^3 \quad \text{(F.D.)} \end{aligned}$$

where $\zeta(3) \simeq 1.2$ comes from doing the integral. The energy density is given by

$$\begin{aligned} \rho c^2 &= g\frac{\pi^2}{30}(k_BT)\left(\frac{k_BT}{\hbar c}\right)^3 \quad \text{(B.E.)} \\ \rho c^2 &= \frac{7}{8}g\frac{\pi^2}{30}(k_BT)\left(\frac{k_BT}{\hbar c}\right)^3 \quad \text{(F.D.)} \end{aligned}$$

and pressure satisfies $P = \rho c^2/3$. As you can see these are the properties of a radiation. In other words, even massive particles will behave like radiation at sufficiently high temperatures. At low temperatures we have $k_BT \ll Mc^2$ and for both fermions and bosons the macroscopic quantities are given by:

$$\begin{aligned} n &= g\left(\frac{2\pi}{h^2}\right)^{\frac{3}{2}}(Mk_BT)^{3/2}\exp(-\frac{Mc^2}{k_BT}) \\ \rho c^2 &= Mc^2n \\ P &= nk_BT \ll Mc^2n = \rho. \end{aligned}$$

This last expression tells us that the pressure is negligible as it should be for non-relativistic matter.

This calculation has already given us an insight into how matter evolves during expansion. At sufficiently early times it all looks like radiation. As it cools down and the temperature falls below mass thresholds, the number of particles behaving relativistically decreases until when we get to today, there are effectively only three type of particles which behave relativistically: the three types of neutrinos. We denote the *effective number of relativistic degrees of freedom* by g_* and the energy density in relativistic degrees of freedom is given by

$$\rho = g_*\frac{\pi^2}{30}(k_BT)\left(\frac{k_BT}{\hbar c}\right)^3$$

22 The Thermal history of the Universe: The Cosmic Microwave Background

Until now we have considered things evolving passively, subjected to the expansion of the Universe. But we know that the interactions between different components of matter can be far more complex. Let us focus on the realm of chemistry, in particular on the interaction between one electron and one proton. From atomic physics and quantum mechanics you already know that an electron and a proton may bind together to form a Hydrogen atom. To tear the electron away we need an energy of about $13.6eV$. But imagine now that the universe is sufficiently hot that there are particles zipping around that can knock the electron out of the atom. We can imagine that at high temperatures it will be very difficult to keep electrons and protons bound together. If the temperature of the Universe is such that $T \simeq 13.6eV$ then we can imagine that there will be a transition between ionized and neutral hydrogen.

We can work this out in more detail (although not completely accurately) if we assume that this transition occurs in thermal equilibrium throughout. Let us go through the steps that lead to the *Saha equation*. Assume we have an equilibrium distribution of protons, electrons and hydrogen atoms. Let n_p, n_e and n_H be their number densities. In thermal equilibrium (with $T \ll M$) we have that the number densities are given by

$$n_i = g_i \left(\frac{2\pi}{h^2}\right)^{\frac{3}{2}} (M_i k_B T)^{\frac{3}{2}} \exp \frac{\mu_i - M_i c^2}{k_B T}$$

where $i = p, n, H$. In chemical equlibrium we have that

$$\mu_p + \mu_e = \mu_H$$

so that

$$n_H = g_H \left(\frac{2\pi}{h^2}\right)^{\frac{3}{2}} (M_H k_B T)^{\frac{3}{2}} \exp \frac{-M_H c^2}{k_B T} \exp \frac{(\mu_p + \mu_e)}{k_B T}$$

We can use the expressions for n_p and n_e to eliminate the chemical potentials and obtain:

$$n_H = n_e n_p \frac{g_H}{g_p g_e} \left(\frac{2\pi}{h^2}\right)^{-\frac{3}{2}} (M_H k_B T)^{\frac{3}{2}} (M_p k_B T)^{-\frac{3}{2}} (M_e k_B T)^{-\frac{3}{2}} \exp \frac{-M_H c^2 + M_p c^2 + M_e c^2}{k_B T}$$

There are a series of simplifications we can now consider: i) $M_p \simeq M_H$, ii) the binding energy is $B \equiv -M_H c^2 + M_p c^2 + M_e c^2 = 13.6eV$, iii) $n_B = n_p + n_H$ iv) $n_e = n_p$ and finally $g_p = g_e = 2$ and $g_H = 4$. So we end up with

$$n_H = n_p^2 (M_e k_B T)^{-\frac{3}{2}} \left(\frac{2\pi}{h^2}\right)^{-\frac{3}{2}} \exp \frac{B}{k_B T}$$

We can go further and define an ionization fraction $X \equiv \frac{n_p}{n_B}$. Quite clearly we have X is 1 if the Universe is completely ionized and 0 if it is neutral. Using the definition of the baryon to entropy fraction we have

$$1 - X = X^2 \eta_B n_\gamma \left(\frac{2\pi}{h^2}\right)^{-\frac{3}{2}} (M_e k_B T)^{-\frac{3}{2}} \exp \frac{B}{k_B T} \tag{25}$$

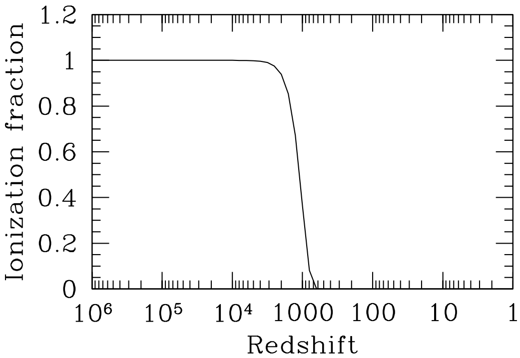

Figure 11: The evolution of the ionization fraction as a function of redshift

Finally we have that we are in thermal equilibrium so we have an expression for n_γ and we get

$$\frac{1-X}{X^2} \simeq 3.8\eta_B \left(\frac{k_B T}{M_e c^2}\right)^{\frac{3}{2}} \exp\frac{B}{k_B T} \tag{26}$$

This is the Saha equation. It tells us how the ionization fraction, X evolves as a function of time. At sufficiently early times we will find that $X = 1$, i.e. the Universe is completely ionized. As it crosses a certain threshold, electrons and protons combine to form Hydrogen. This happens when the temperature of the Universe is $T \simeq 3570K$ or $0.308eV$, i.e. when it was approximately $380,000$ years old, at a redshift of $z \simeq 1100$. We would naively expect this to happen at $13.6eV$ but the prefactors in front of the exponential play an important role. One way to think about it is that, at a given temperature there will always be a few photons with energies larger than the average temperature. Thus energetic photons only become unimportant at sufficiently low temperatures.

What does this radiation look like to us? At very early times, before recombination, this radiation will be in thermal equilibrium and satisfy the Planck spectrum:

$$\rho(\nu)d\nu = \frac{8\pi h}{c^3}\frac{\nu^3 d\nu}{\exp(h\nu/k_B T)-1}$$

After recombination, the electrons and protons combine to form neutral hydrogen and the photons will be left to propagate freely. The only effect will be the redshifting due to the expansion. The net effect is that the shape of the spectrum remains the same, the peak shifting as $T \propto 1/a$. So even though the photons are not in thermal equilibrium anymore, the spectrum will still be that of thermal equilibrium with the temperature $T_0 = 3000^{\circ}/1100$ Kelvin, i.e. $T_0 = 2.75^{\circ}$ Kelvin.

The history of each individual photon can also be easily described. Let's work backwards. After recombination, a photon does not interact with anything and simply propagates forward at the speed of light. It's path will be a straight line starting off at the time of recombination and ending today. Before recombination, photons are highly interacting with a very dense medium of charged particles, the protons and electrons. This means that they are constantly scattering off particles, performing something akin to a random walk with a very small step length. For all intents and purposes, they are glued to the spot unable to move. So one can think of such a photon's history as starting off stuck at some point in space and, at recombination, being released to propogate forward until now.

We can take this even further. If we look from a specific observing point (such as the Earth or a satellite), we will be receiving photons from all directions that have been travelling in a straight line since the Universe recombined. All these straight lines will have started off at the same time and at the same distance from us-i.e. they will have started off from the surface of a sphere. This surface, known as the *surface of last scattering* is what we see when we look at the relic radiation. It is very much like a photograph of the Universe when it was 380,000 years old.

23 The Thermal history of the Universe: out of equilibirium and Big Bang Nucleosynthesis

We have assumed that the Universe is in thermal equilibrium throughout this process. We have come up with an expression for the ionization fraction which is not completely accurate but qualitatively has the correct behaviour. There is another situation where assuming thermal equilibrium will not only give us the wrong quantitative but also the wrong qualitative result. We shall now look at what happens when the temperature of the Universe is $k_BT \simeq 1MeV$. This corresponds to the energy where nuclear processes play an important role.

The particles we will consider are protons (p) and neutrons (n). These particles will combine to form the nuclei of the elements. For example

$$\begin{aligned} 1p &\rightarrow \text{Hydrogen nucleus} \\ 1p+1n &\rightarrow \text{Deuterium nucleus} \\ 1p+2n &\rightarrow \text{Tritium nucleus} \\ 2p+1n &\rightarrow \text{Helium 3 nucleus} \\ 2p+2n &\rightarrow \text{Helium 4 nucleus} \end{aligned} \tag{27}$$

The binding energy of each nucleus will be the difference between the mass of the nucleus and the sum of masses of the protons and neutrons that form it. For example the binding energy of Deuterium is $B_D = 2.22MeV$. The neutrons and protons can convert into each other through the weak interactions:

$$\begin{aligned} p+\bar{\nu}_e &\rightarrow n+e^+ \\ p+e^- &\rightarrow n+\nu \\ n &\rightarrow p+e^-+\bar{\nu} \end{aligned} \tag{28}$$

Let us start with an equilibrium approach. We can try to work out the abundance of light elements of the Universe using the same rationale we used above. Once again we have

$$n_i = g_i\left(\frac{2\pi}{h^2}\right)^{\frac{3}{2}}(M_ik_BT)^{\frac{3}{2}}\exp\frac{\mu_i - M_ic^2}{k_BT}$$

Let us assume thermal equilibrium ($\mu_n = \mu_p$). We have that

$$\frac{n_n^{eq}}{n_p^{eq}} = \exp\left(-\frac{Q}{k_B T}\right) \tag{29}$$

where Q is the energy due to the mass difference between the neutron and the proton, $Q = 1.293 MeV$. We can see that at high temperatures there are as many protons as neutrons. But as T falls below Q the mass difference becomes important and the neutrons dwindle away. If this were the correct way of calculating the abundance of neutrons we would find that as the Universe cools, all the neutrons would dissapear. No neutrons would be left.

To get an accurate estimate we must go beyond the equilibrium approximation. We can step back a bit and think about what is actually going on when protons and neutrons are interconverting into each other. The reaction can be characterised in terms of a reaction rate, Γ which has units of s^{-1}. The reaction must compete against the expansion of the universe which itself can be described in terms of a "rate": the expansion rate H which has units of s^{-1}. The relative sizes of Γ and H dictate how important the reactions are in keeping the neutrons and protons equilibrated.

One can write down a Boltzmann equation for the comoving neutron number (the number of neutrons in a box in comoving units):

$$\frac{d\ln N_n}{d\ln a} = -\frac{\Gamma}{H}\left[1 - \left(\frac{N_n^{eq}}{N_n}\right)^2\right]$$

Where $N_n^{eq} = a^3 n_n^{eq}$ is the equilibrium expression given above. If $\Gamma \gg H$ we have that $N_n \to N_n^{eq}$, i.e. the neutron number density will be pushed to its equilibrium value. In that regime we will have the ratio of neutrons as given by 29. If, however, $\Gamma \ll H$, the expansion of the universe will win out and inhibit the depletion or creation of neutrons through that reaction. The equation is then approximately given by

$$\frac{d\ln N_n}{d\ln a} \simeq 0$$

i.e. the neutron comoving number is frozen (and the number density will decay as a^{-3}). The transition from one regime to the other will occur when $\Gamma \sim H$ and it will depend on how the reaction rate depends on temperature and masses. It turns out that for this reaction, the temperature at which reactions "freeze out" is $k_B T_f \simeq 0.7 - 0.8 MeV$. The relative number density of neutrons to protons will be frozen in at

$$\frac{n_n^{eq}}{n_p^{eq}} \simeq \frac{1}{6}$$

In fact the neutron decay rate plays a role as well to further decrease the fraction of neutrons so that we in fact get $n_n/n_p \to \frac{1}{7}$. We can use a very simple argument to find the fraction of Helium 4 in the Universe. We start off with 7/8 in protons and 1/8 in neutrons. Let us assume that the neutrons are used to make Helium atoms. We then need to pair up the 1/8 with 1/8 protons, reducing the number of unpaired protons to $6/8 \simeq 75\%$. So we roughly expect to have about 25% of the mass in Helium and 75% in Hydrogen. A more accurate calculation gives a

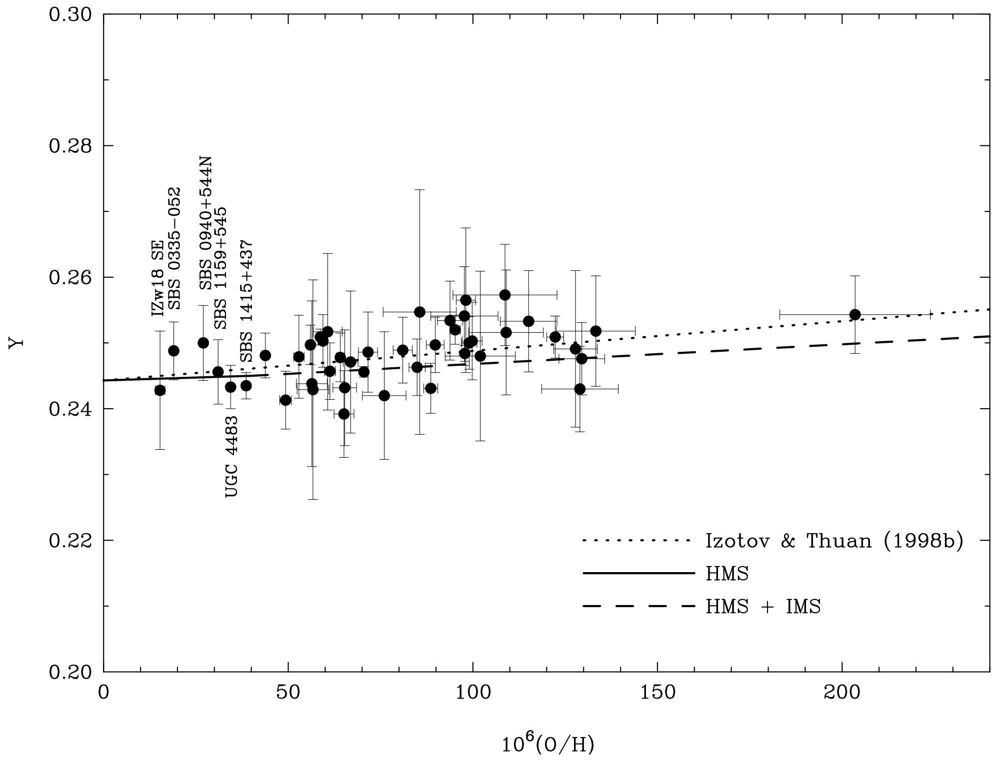

Figure 12: The mass fraction of Helium in stars as a function of the fraction of Oxygen (Izotov & Thuan). The more Oxygen there is, there more stellar burning has occured and therefore more helium has been produced from hydrogen burning. This means that we expect to find a higher amount of helium than what we have found from primordial nucleosynthesis

helium fraction of about 24% which is borne out by observations. One can look at astrophysical systems and measure the amount of Helium. The more Oxygen there is, the more processing there is (and hence the more Helium has been produced in stellar burning).

The abundances of the other light elements is very small but measurable and predictable. For example

$$\frac{n_D}{n_H} \simeq 3 \times 10^{-5}$$

Throughout the history of the Universe, relics have been left over. A relic bath of photons is left over from when the Universe had a $k_B T \simeq eV$. A relic distribution of light elements is left over from when $k_B T \simeq MeV$. It is conceivable that relic particles are left over from transitions which may have occured at higher energies and temperatures.